

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/03790

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : 435/4, 91.2, 320.1; 530/350, 387.1; 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/4, 91.2, 320.1; 530/350, 387.1; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN, MEDLINE, BIOSIS

search terms: macular dystrophy, protein, enzyme, assay, activation, stimulation, agonist, antagonist, inhibit, screen

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,705,380 A (NORTH et al) 06 January 1998, abstract col. 9, lines 39-54.	18
X	US 5,686,598 A (NORTH et al) 11 November 1997, abstract, col. 9, lines 39-54.	18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be of particular relevance

B earlier document published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

Z

document member of the same patent family

Date of the actual completion of the international search

26 MAY 1999

Date of mailing of the international search report

16 JUN 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

HOLLY SCHNIZER

Telephone No.

(703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/03790

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-17
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Claims 1-17 are directed to DNA molecules and proteins having specific nucleotide and amino acid sequences not in compliance with the PCT sequence listing requirements. Therefore, the claims could not be understood well enough to permit searching.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/03790

A. CLASSIFICATION OF SUBJECT MATTER: IPC (6):

C07H 21/02; C07K 16/00, 14/00; C12N 15/63; C12P 19/34; C12Q 1/00

PCT

**NOTIFICATION OF THE RECORDING
OF A CHANGE**

(PCT Rule 92bis.1 and
Administrative Instructions, Section 422)

From the INTERNATIONAL BUREAU

To:

MERCK & CO., INC.
126 East Lincoln Avenue
Rahway, NJ 07065
ETATS-UNIS D'AMERIQUE

Date of mailing (day/month/year)
13 July 2000 (13.07.00)

Applicant's or agent's file reference
PCT 20177Y

International application No.
PCT/US99/03790

IMPORTANT NOTIFICATION

International filing date (day/month/year)
22 February 1999 (22.02.99)

1. The following indications appeared on record concerning:

☒ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality
SE

State of Residence
SE

Telephone No.

Facsimile No.

Teleprinter No.

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☒ the name ☒ the address ☐ the nationality ☐ the residence

Name and Address

WADELIUS, Claes
P.O. Box 256
S-751 05 Uppsala
Sweden

State of Nationality
SE

State of Residence
SE

Telephone No.

Facsimile No.

Teleprinter No.

3. Further observations, if necessary:

Additional inventor and applicant for the US only.

4. A copy of this notification has been sent to:

☒ the receiving Office ☐ the designated Offices concerned
☐ the International Searching Authority ☒ the elected Offices concerned
☐ the International Preliminary Examining Authority ☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Philippe Bécamel

Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92bis.1 and
Administrative Instructions, Section 422)

From the INTERNATIONAL BUREAU

To:

MERCK & CO., INC.
126 East Lincoln Avenue
Rahway, NJ 07065
ETATS-UNIS D'AMERIQUE

Date of mailing (day/month/year) 14 July 2000 (14.07.00)	
Applicant's or agent's file reference PCT 20177Y	IMPORTANT NOTIFICATION
International application No. PCT/US99/03790	International filing date (day/month/year) 22 February 1999 (22.02.99)

1. The following indications appeared on record concerning:

☒ the applicant
 ☒ the inventor
 ☐ the agent
 ☐ the common representative

Name and Address 	State of Nationality SE	State of Residence SE
	Telephone No.	
	Facsimile No.	
	Teleprinter No.	

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person
 ☒ the name
 ☒ the address
 ☐ the nationality
 ☐ the residence

Name and Address WADELIUS, Claes P.O. Box 256 S-751 05 Uppsala Sweden	State of Nationality SE	State of Residence SE
	Telephone No.	
	Facsimile No.	
	Teleprinter No.	

3. Further observations, if necessary:

CORRECTED VERSION OF PCT/IB/306 FORM SENT ON 13 JULY 2000. WADELIUS, Claes is inventor and applicant for all the designated States.

4. A copy of this notification has been sent to:

<input checked="" type="checkbox"/> the receiving Office	<input type="checkbox"/> the designated Offices concerned
<input type="checkbox"/> the International Searching Authority	<input checked="" type="checkbox"/> the elected Offices concerned
<input type="checkbox"/> the International Preliminary Examining Authority	<input type="checkbox"/> other:

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No.: (41-22) 740.14.35	Authorized officer Philippe Bécamel Telephone No.: (41-22) 338.83.38
---	---

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark
Office
Box PCT
Washington, D.C. 20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year)
20 October 1999 (20.10.99)

International application No.
PCT/US99/03790

Applicant's or agent's file reference
PCT 20177Y

International filing date (day/month/year)
22 February 1999 (22.02.99)

Priority date (day/month/year)
25 February 1998 (25.02.98)

Applicant

PETRUKHIN, Konstantin et al

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:
22 September 1999 (22.09.99)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was

☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

R. Forax

Telephone No.: (41-22) 338.83.38

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

REC'D 03 JUL 2000

WIPO

PCT

Applicant's or agent's file reference 20177Y-PCT	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/US99/03790	International filing date (day/month/year) 22 FEBRUARY 1999	Priority date (day/month/year) 25 FEBRUARY 1998
International Patent Classification (IPC) or national classification and IPC Please See Supplemental Sheet.		
Applicant MERCK & CO., INC.		

- This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.
- This REPORT consists of a total of 5 sheets.
☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority. (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).
These annexes consist of a total of 0 sheets.

- This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☒ Non-establishment of report with regard to novelty, inventive step or industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand 22 SEPTEMBER 1999	Date of completion of this report 07 JUNE 2000
Name and mailing address of the IPEA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer JOYCE BRIDGERS PARALEGAL SPECIALIST CHEMICAL MATRIX HOLLY SCHNIZER <i>[Signature]</i>
Facsimile No. (703) 305-3230	Telephone No. (703) 308-4028

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US99/03790

I. Basis of the report

1. With regard to the elements of the international application:*

☒ the international application as originally filed

☒ the description:

pages 1-41

pages NONE, as originally filed

pages NONE, filed with the demand

☒ the claims: 42-44

pages

pages NONE, as originally filed

pages NONE, as amended (together with any statement) under Article 19

pages NONE, filed with the demand

☒ the drawings: 1-18

pages

pages NONE, as originally filed

pages NONE, filed with the demand

☒ the sequence listing part of the description:

pages NONE

pages NONE, as originally filed

pages NONE, filed with the demand

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language _____ which is:

☐ the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).

☐ the language of publication of the international application (under Rule 48.3(b)).

☐ the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).

3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

☐ contained in the international application in printed form.

☐ filed together with the international application in computer readable form.

☐ furnished subsequently to this Authority in written form.

☐ furnished subsequently to this Authority in computer readable form.

☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. ☒ The amendments have resulted in the cancellation of:

☒ the description, pages none

☒ the claims, Nos. none

☒ the drawings, sheets/fig none

5. ☒ This report has been drawn as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

* Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17).

**Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US99/03790

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been and will not be examined in respect of:

☐ the entire international application.

☒ claims Nos. 1-17

because:

☐ the said international application, or the said claim Nos. _ relate to the following subject matter which does not require international preliminary examination (*specify*).

☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. _ are so unclear that no meaningful opinion could be formed (*specify*).

☐ the claims, or said claims Nos. _ are so inadequately supported by the description that no meaningful opinion could be formed.

☒ no international search report has been established for said claims Nos. 1-17.

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

☐ the written form has not been furnished or does not comply with the standard.

☐ the computer readable form has not been furnished or does not comply with the standard.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US99/03790

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**1. statement**

Novelty (N)

Claims	<u>none</u>	YES
Claims	<u>18</u>	NO

Inventive Step (IS)

Claims	<u>none</u>	YES
Claims	<u>18</u>	NO

Industrial Applicability (IA)

Claims	<u>18</u>	YES
Claims	<u>none</u>	NO

2. citations and explanations (Rule 70.7)

Claim 18 novelty under PCT Article 33(2) as being anticipated by North et al. (U.S. Pat. No. 5,705,380).

The North et al. patent describes a gene responsible for an autosomal dominant con-rod retinal dystrophy and methods for screening for substances which modulate the function of the encoded protein. Therefore, the North et al. reference anticipates Claim 18.

Claim 18 lacks novelty under PCT Article 33(2) as being anticipated by North et al. (U.S. Pat. No. 5,686,598, 1997).

The North et al. patent describes the gene responsible for an autosomal dominant con-rod retinal dystrophy and methods for screening substances which modulate the function of the encoded protein. Therefore, the North et al. reference anticipates Claim 18.

Claim 18 has industrial applicability because the claimed method could be used to find compounds which potentially could be used in the treatment of Macular Dystrophy.

----- NEW CITATIONS -----

NONE

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US99/03790

Supplemental Box

(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: Boxes I - VIII

Sheet 10

CLASSIFICATION:

The International Patent Classification (IPC) and/or the National classification are as listed below:
IPC(6): C07H 21/02; C07K 16/00, 14/00; C12N 15/63; C12P 19/34; C12Q 1/00 and US Cl.: 435/4, 91.2, 320.1; 530/350, 387.1; 536/23.1

I. BASIS OF REPORT:

5. (Some) amendments are considered to go beyond the disclosure as filed:
NONE

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07H 21/02, C07K 16/00, 14/00, C12N 15/63, C12P 19/34, C12Q 1/00	A1	(11) International Publication Number: WO 99/43695 (43) International Publication Date: 2 September 1999 (02.09.99)
(21) International Application Number: PCT/US99/03790 (22) International Filing Date: 22 February 1999 (22.02.99) (30) Priority Data: 60/075,941 25 February 1998 (25.02.98) US 60/112,926 18 December 1998 (18.12.98) US (71) Applicants (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). UPPSALA UNIVERSITY [SE/SE]; P.O. Box 256, S-751 05 Uppsala (SE). (72) Inventors; and (75) Inventors/Applicants (for US only): PETRUKHIN, Konstantin [RU/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US)/CASKEY, C., Thomas [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). METZKER, Michael [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). WADELIUS, Claes [SE/SE]; P.O. Box 256, S-751 05 Uppsala (SE). (74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: BEST'S MACULAR DYSTROPHY GENE		
(57) Abstract Novel human and mouse DNA sequences that encode the gene CG1CE, which, when mutated, is responsible for Best's macular dystrophy, are provided. Provided are genomic CG1CE DNA as well as cDNA that encodes the CG1CE protein. Also provided is CG1CE protein encoded by the novel DNA sequences. Methods of expressing CG1CE protein in recombinant systems are provided. Also provided are diagnostic methods that detect patients having mutant CG1CE genes.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KC	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE OF THE INVENTION
BEST'S MACULAR DYSTROPHY GENE

CROSS-REFERENCE TO RELATED APPLICATIONS
5 Not applicable.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D
Not applicable.

10 REFERENCE TO MICROFICHE APPENDIX
Not applicable.

FIELD OF THE INVENTION

15 The present invention is directed to novel human and mouse DNA sequences encoding a protein which, when present in mutated form, results in the occurrence of Best's Macular Dystrophy.

BACKGROUND OF THE INVENTION

20 Macular dystrophy is a term applied to a heterogeneous group of diseases that collectively are the cause of severe visual loss in a large number of people. A common characteristic of macular dystrophy is a progressive loss of central vision resulting from the degeneration of the pigmented epithelium underlying the retinal macula. In many forms of macular dystrophy, the end stage of the disease results in legal
25 blindness. More than 20 types of macular dystrophy are known: e.g., age-related macular dystrophy, Stargardt's disease, atypical vitelliform macular dystrophy (VMD1), Usher Syndrome Type 1B, autosomal dominant neovascular inflammatory vitreoretinopathy, familial exudative vitreoretinopathy, and Best's macular dystrophy (also known
30 as hereditary macular dystrophy or Best's vitelliform macular dystrophy (VMD2)). For a review of the macular dystrophies, see Sullivan & Daiger, 1996, Mol. Med. Today 2:380-386.

35 Best's Macular Dystrophy (BMD) is an inherited autosomal dominant macular dystrophy of unknown biochemical cause. BMD has an age of onset that can range from childhood to after 40. Clinical symptoms include, at early stages, an abnormal accumulation of the

yellowish material lipofuscin in the retinal pigmented epithelium (RPE) underlying the macula. This gives rise to a characteristic "egg yolk" appearance of the RPE and gradual loss of visual acuity. With increasing age, the RPE becomes more and more disorganized, as the lipofuscin accumulations disperse and scarring and neovascularization take place. These changes are accompanied by further loss of vision.

The pathological features seen in BMD are in many ways similar to the features seen in age-related macular dystrophy, the leading cause of blindness in older patients in the developed world. Age-related macular dystrophy is an extraordinarily difficult disease to study genetically, since by the time patients are diagnosed, their parents are usually no longer living and their children are still asymptomatic. Thus, family studies which have led to the discovery of the genetic basis of many other diseases have not been practical for age-related macular dystrophy. As there are currently no widely effective treatments for age-related macular dystrophy, it is hoped that study of BMD, and in particular the discovery of the underlying genetic cause of BMD, will shed light on age-related macular dystrophy as well.

Linkage analysis has established that the gene responsible for BMD resides in the pericentric region of chromosome 11, at 11q13, near the markers D11S956, FCER1B, and UGB (Forsman et al., 1992, Clin. Genet. 42:156-159; Hou et al., 1996, Human Heredity 46:211-220). Recently, the gene responsible for BMD was localized to a ~1.7 mB PAC contig lying mostly between the markers D11S1765 and UGB (Cooper et al., 1997, Genomics 41:185-192). Recombination breakpoint mapping in a large Swedish pedigree limited the minimum genetic region containing the BMD gene to a 980 kb interval flanked by the microsatellite markers D11S4076 and UGB (Graff et al., 1997, Hum. Genet. 101: 263-279).

One difficulty in diagnosing BMD is that carriers of the diseased gene for BMD may be asymptomatic in terms of visual acuity and morphological changes of the RPE observable in a routine ophthalmologic examination. There does exist a test, the electro-oculographic examination (EOG), which detects differences in electrical potential between the cornea and the retina, that can distinguish asymptomatic BMD patients from normal individuals. However, the EOG requires specialized, expensive equipment, is difficult to

administer, and requires that the patient be present at the site of the equipment when the test is performed. It would be valuable to have an alternative method of diagnosing asymptomatic carriers of mutations in the gene responsible for BMD that is simpler, less expensive, and does not require the presence of the patient while the test is being performed. For example, a diagnostic test that relies on a blood sample from a patient suspected of being an asymptomatic carrier of BMD would be ideal.

SUMMARY OF THE INVENTION

The present invention is directed to novel human and mouse DNA sequences that encode the gene CG1CE, which, when mutated, is responsible for Best's macular dystrophy. The present invention includes genomic CG1CE DNA as well as cDNA that encodes the CG1CE protein. The human genomic CG1CE DNA is substantially free from other nucleic acids and has the nucleotide sequence shown in SEQ.ID.NO.:1. The human cDNA encoding CG1CE protein is substantially free from other nucleic acids and has the nucleotide sequence shown in SEQ.ID.NO.:2 or SEQ.ID.NO.:4. The mouse cDNA encoding CG1CE protein is substantially free from other nucleic acids and has the nucleotide sequence shown in SEQ.ID.NO.:28. Also provided is CG1CE protein encoded by the novel DNA sequences. The human CG1CE protein is substantially free from other proteins and has the amino acid sequence shown in SEQ.ID.NO.:3 or SEQ.ID.NO.:5. The mouse CG1CE protein is substantially free from other proteins and has the amino acid sequence shown in SEQ.ID.NO.:29. Methods of expressing CG1CE protein in recombinant systems are provided. Also provided are diagnostic methods that detect carriers of mutant CG1CE genes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A-F shows the genomic DNA sequence of human CG1CE (SEQ.ID.NO.:1). Underlined nucleotides in capitals represent exons. The start ATG codon in exon 2 and the stop TAA codon in exon 11 are shown in bold italics. The consensus polyadenylation signal AATAAA in exon 11 is shown in bold. The alternatively spliced part of

exon 7 is shown in underlined italics. The exact lengths of two gaps between exons 1 and 2 and between exons 7 and 8 are unknown; these gaps are presented as runs of ten Ns for the sake of convenience. The portion of exon 11 beginning at position 15,788 represents the 3' untranslated region; 132 base pairs downstream of the polyadenylation signal of the CG1CE gene are multiple ESTs, representing the 3' untranslated region of the ferritin heavy chain gene (FTH). FTH has been mapped to human chromosome 11q13 (Hentze *et al.*, 1986, Proc. Nat. Acad. Sci. 83: 7226-7230); the FTH gene was later shown to be a part of the smallest minimum genetic region containing the BMD gene, as determined by recombination breakpoint mapping in a 12 generation Swedish pedigree (Graff *et al.*, 1997, Hum. Genet. 101: 263-279).

Figure 2 shows the complete sequence of the short form of human CG1CE cDNA (SEQ.ID.NO.:2). The ATG start codon is at position 105; the TAA stop codon is at position 1,860.

Figure 3 shows the complete amino acid sequence of the long form of human CG1CE protein (SEQ.ID.NO.:3). This long form of the human CG1CE protein is produced by translation of the short form of CG1CE cDNA.

Figure 4 shows the complete sequence of the long form of human CG1CE cDNA (SEQ.ID.NO.:4). This long form of the human CG1CE cDNA is produced when an alternative splice donor site is utilized in intron 7. The ATG start codon is at position 105; the TGA stop codon is at position 1410.

Figure 5 shows the complete amino acid sequence of the short form of the human CG1CE protein (SEQ.ID.NO.:5). This short form of the human CG1CE protein is produced by translation of the long form of CG1CE cDNA.

Figure 6 shows the results of sequencing runs of PCR fragments that represent exon 4 and adjacent intronic regions from three individuals from the Swedish pedigree S1, two of whom are affected with BMD. From top to bottom, the runs are: patient S1-5 (homozygous affected with BMD), sense orientation; patient S1-4 (heterozygous affected with BMD), sense orientation; patient S1-3 (normal control, unaffected sister of S1-4), sense orientation; patient S1-5 (affected with BMD), anti-sense orientation; patient S1-4 (affected with

BMD), ~~anti~~-sense orientation; patient S1-3 (normal control), anti-sense orientation. Reading from left to right, the mutation shows up at position 31 of the sequence shown in the case of patients S1-5 and S1-4. The mutation in family S1 changes tryptophan to cysteine.

5 Figure 7 shows a multiple sequence alignment of human CG1CE protein with partial sequences of related proteins from *C. elegans*. Related proteins from *C. elegans* were identified by BLASTP analysis of non-redundant GenBank database. This figure shows that
10 two amino acids mutated in two different Swedish families with BMD (families S1 and SL76) are evolutionarily conserved. 15 of 16 related proteins from *C. elegans* contain a tryptophan at the position of the mutation in family S1, as does the wild-type CG1CE gene. Only one *C. elegans* protein does not have a tryptophan at the position of the
15 mutation. In this protein (accession number p34577), tryptophan is changed for isofunctional phenylalanine (phenylalanine is highly similar to tryptophan in that it also is a hydrophobic aromatic amino acid). Mutation in the BMD family SL76 changes a tyrosine to histidine. Again, all 16 related proteins from *C. elegans* contain tyrosine or
20 isofunctional phenylalanine in this position (tyrosine is highly similar to phenylalanine in that it also is an aromatic amino acid).

Figure 8A-C shows the complete sequence of mouse CG1CE cDNA (SEQ.ID.NO.:28) and mouse CG1CE protein (SEQ.ID.NO.:29).

Figure 9A-B shows an alignment of the amino acid sequences of the long form of human CG1CE protein (SEQ.ID.NO.:3) and
25 mouse CG1CE protein (SEQ.ID.NO.:29). In this figure, CG1CE is referred to as "bestrophin."

Figure 10A-C shows the results of *in situ* hybridization experiments demonstrating that mouse CG1CE mRNA expression is localized to the retinal pigmented epithelium cells (RPE). Figure 10A
30 shows the results of using an antisense CG1CE probe. The antisense probe hybridizes to mouse CG1CE mRNA present in the various cell layers of the retina, labeling with dark bands the cells containing CG1CE mRNA. The antisense probe strongly hybridized to the RPE cells and not to the cells of the other layers of the retina. Figure 10B shows
35 the results using a sense CG1CE probe as a control. The sense probe does not hybridize to CG1CE mRNA and does not label the RPE cells.

Figure 10C is a higher magnification of the RPE cells from Figure 10A. Human CG1CE mRNA shows a similar distribution, being confined to the RPE cells of the human retina.

5 DETAILED DESCRIPTION OF THE INVENTION

For the purposes of this invention:

“Substantially free from other proteins” means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other proteins. Thus, a CG1CE protein preparation that is
10 substantially free from other proteins will contain, as a percent of its total protein, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non- CG1CE proteins. Whether a given CG1CE protein
15 preparation is substantially free from other proteins can be determined by such conventional techniques of assessing protein purity as, *e.g.*, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) combined with appropriate detection methods, *e.g.*, silver staining or immunoblotting.

“Substantially free from other nucleic acids” means at least
20 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other nucleic acids. Thus, a CG1CE DNA preparation that is substantially free from other nucleic acids will contain, as a percent of its total nucleic acid, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more
25 than 0.1%, of non- CG1CE nucleic acids. Whether a given CG1CE DNA preparation is substantially free from other nucleic acids can be determined by such conventional techniques of assessing nucleic acid purity as, *e.g.*, agarose gel electrophoresis combined with appropriate staining methods, *e.g.*, ethidium bromide staining, or by sequencing.

30 A “conservative amino acid substitution” refers to the replacement of one amino acid residue by another, chemically similar, amino acid residue. Examples of such conservative substitutions are: substitution of one hydrophobic residue (isoleucine, leucine, valine, or methionine) for another; substitution of one polar residue for another
35 polar residue of the same charge (*e.g.*, arginine for lysine; glutamic acid

for aspartic acid); substitution of one aromatic amino acid (tryptophan, tyrosine, or phenylalanine) for another.

The present invention relates to the identification and cloning of CG1CE, a gene which, when mutated, is responsible for Best's macular dystrophy. That CG1CE is the Best's macular dystrophy gene is supported by various observations:

1. CG1CE maps to the genetically defined region of human chromosome 11q12-q13 that has been shown to contain the Best's macular dystrophy gene. CG1CE is present on two PAC clones, 759J12 and 466A11, that lie precisely in the most narrowly defined region that has been shown to contain CG1CE (Cooper *et al.*, 1997, Genomics 41:185-192; Stühr *et al.*, 1997, Genome Res. 8:48-56; Graff *et al.*, 1997, Hum. Genet. 101: 263-279).

2. CG1CE is expressed predominately in the retina.

3. In patients having Best's macular dystrophy, CG1CE contains mutations in evolutionarily conserved amino acids.

4. The CG1CE genomic clones contain another gene (FTH) that has been physically associated with the Best's macular dystrophy region (Cooper *et al.*, 1997, Genomics 41:185-192; Stühr *et al.*, 1997, Genome Res. 8:48-56; Graff *et al.*, 1997, Hum. Genet. 101:263-279). The FTH and CG1CE genes are oriented tail-to-tail; the distance between their polyadenylation signals is 132 bp.

The present invention provides DNA encoding CG1CE that is substantially free from other nucleic acids. The present invention also provides recombinant DNA molecules encoding CG1CE. The present invention provides DNA molecules substantially free from other nucleic acids comprising the nucleotide sequence shown in Figure 1 as SEQ.ID.NO.:1. Analysis of SEQ.ID.NO.:1 revealed that this genomic sequence defines a gene having 11 exons. These exons collectively have an open reading frame that encodes a protein of 585 amino acids. If an alternative splice donor site is utilized in exon 7, a cDNA containing an additional 203 bases is produced. Although longer, this cDNA contains a shorter open reading frame of 1,305 bases (due to the presence of a change in reading frame that introduces a stop codon) that encodes a protein of 435 amino acids. Thus, the present invention includes two cDNA molecules encoding two forms of CG1CE protein that are

substantially free from other nucleic acids and have the nucleotide sequences shown in Figure 2 as SEQ.ID.NO.:2 and in Figure 4 as SEQ.ID.NO.:4.

5 The present invention includes DNA molecules substantially free from other nucleic acids comprising the coding regions of SEQ.ID.NO.:2 and SEQ.ID.NO.:4. Accordingly, the present invention includes DNA molecules substantially free from other nucleic acids having a sequence comprising positions 105-1,859 of SEQ.ID.NO.:2 and positions 105-1,409 of SEQ.ID.NO.:4. Also included are recombinant
10 DNA molecules having a nucleotide sequence comprising positions 105-1,859 of SEQ.ID.NO.:2 and positions 105-1,409 of SEQ.ID.NO.:4.

Portions of the cDNA sequences of SEQ.ID.NO.:2 and SEQ.ID.NO.:4 are found in two retina-specific ESTs deposited in GenBank by The Institute for Genomic Research (accession numbers
15 AA318352 and AA317489). Other ESTs that correspond to this cDNA are accession numbers AA307119 (from a colon carcinoma), AA205892 (from neuronal cell line), and AA326727 (from human cerebellum). A true mouse ortholog of the CG1CE gene is represented in the mouse EST
AA497726 (from mouse testis).

20 The novel DNA sequences of the present invention encoding CG1CE, in whole or in part, can be linked with other DNA sequences, *i.e.*, DNA sequences to which CG1CE is not naturally linked, to form "recombinant DNA molecules" encoding CG1CE. Such other sequences can include DNA sequences that control transcription or translation
25 such as, *e.g.*, translation initiation sequences, promoters for RNA polymerase II, transcription or translation termination sequences, enhancer sequences, sequences that control replication in microorganisms, sequences that confer antibiotic resistance, or sequences that encode a polypeptide "tag" such as, *e.g.*, a polyhistidine
30 tract or the myc epitope. The novel DNA sequences of the present invention can be inserted into vectors such as plasmids, cosmids, viral vectors, P1 artificial chromosomes, or yeast artificial chromosomes.

Included in the present invention are DNA sequences that hybridize to at least one of SEQ.ID.NO.s.:1, 2, or 4 under stringent
35 conditions. By way of example, and not limitation, a procedure using conditions of high stringency is as follows: Prehybridization of filters

containing DNA is carried out for 2 hr. to overnight at 65°C in buffer composed of 6X SSC, 5X Denhardt's solution, and 100 µg/ml denatured salmon sperm DNA. Filters are hybridized for 12 to 48 hrs at 65°C in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Washing of filters is done at 37°C for 1 hr in a solution containing 2X SSC, 0.1% SDS. This is followed by a wash in 0.1X SSC, 0.1% SDS at 50°C for 45 min. before autoradiography.

Other procedures using conditions of high stringency would include either a hybridization carried out in 5XSSC, 5X Denhardt's solution, 50% formamide at 42°C for 12 to 48 hours or a washing step carried out in 0.2X SSPE, 0.2% SDS at 65°C for 30 to 60 minutes.

Reagents mentioned in the foregoing procedures for carrying out high stringency hybridization are well known in the art. Details of the composition of these reagents can be found in, *e.g.*, Sambrook, Fritsch, and Maniatis, 1989, Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Harbor Laboratory Press. In addition to the foregoing, other conditions of high stringency which may be used are well known in the art.

The degeneracy of the genetic code is such that, for all but two amino acids, more than a single codon encodes a particular amino acid. This allows for the construction of synthetic DNA that encodes the CG1CE protein where the nucleotide sequence of the synthetic DNA differs significantly from the nucleotide sequences of SEQ.ID.NOs.:2 or 4, but still encodes the same CG1CE protein as SEQ.ID.NOs.:2 or 4. Such synthetic DNAs are intended to be within the scope of the present invention.

Mutated forms of SEQ.ID.NOs.:1, 2, or 4 are intended to be within the scope of the present invention. In particular, mutated forms of SEQ.ID.NOs.:1, 2, or 4 which give rise to Best's macular dystrophy are within the scope of the present invention. Accordingly, the present invention includes a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 7,259 of SEQ.ID.NO.:1 is T, A, or C rather than G, so that the codon at positions

7,257-7,259 encodes either cysteine or is a stop codon rather than encoding tryptophan. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that at least one of the nucleotides at position 7,257 or 7,258 has been changed so that the codon at positions 7,257-7,259 does not encode tryptophan.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that the nucleotide at position 383 is T, A, or C rather than G, so that the codon at positions 381-383 encodes either cysteine or is a stop codon rather than encoding tryptophan. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that at least one of the nucleotides at position 381 or 382 has been changed so that the codon at positions 381-383 does not encode tryptophan.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that the nucleotide at position 383 is T, A, or C rather than G, so that the codon at positions 381-383 encodes either cysteine or is a stop codon rather than encoding tryptophan. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that at least one of the nucleotides at position 381 or 382 has been changed so that the codon at positions 381-383 does not encode tryptophan.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 7,233 of SEQ.ID.NO.:1 is C, A, or G rather than T, so that the codon at positions 7,233-7,235 does not encode tyrosine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that at least one of the nucleotides at position 7,234 or 7,235 has been changed so that the codon at positions 7,233-7,235 does not encode tyrosine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of

SEQ.ID.NO.:2 except that the nucleotide at position 357 is C, A, or G rather than T, so that the codon at positions 357-359 does not encode tyrosine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of
5 SEQ.ID.NO.:2 except that at least one of the nucleotides at position 358 or 359 has been changed so that the codon at positions 357-359 does not encode tyrosine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of
10 SEQ.ID.NO.:4 except that the nucleotide at position 357 is C, A, or G rather than T, so that the codon at positions 357-359 does not encode tyrosine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of
15 SEQ.ID.NO.:4 except that at least one of the nucleotides at position 358 or 359 has been changed so that the codon at positions 357-359 does not encode tyrosine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 3,330 is C rather than A. Also included in the
20 present invention is a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 3,330 of SEQ.ID.NO.:1 is G, C, or T rather than A, so that the codon at positions 3,330-3,332 does not encode threonine. Also included in the present
invention is a DNA molecule having a nucleotide sequence that is
25 identical to SEQ.ID.NO.:1 except that at least one of the nucleotides at position 3,330 or 3,331 has been changed so that the codon at positions 3,330-3,332 does not encode threonine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of
30 SEQ.ID.NO.:2 except that the nucleotide at position 120 is C rather than A. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of
SEQ.ID.NO.:2 except that the nucleotide at position 120 is G, C, or T
rather than A, so that the codon at positions 120-122 does not encode
35 threonine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of

SEQ.ID.NO.:2 except that at least one of the nucleotides at position 120 or 121 has been changed so that the codon at positions 120-122 does not encode threonine.

5 The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that the nucleotide at position 120 is C rather than A. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of
10 SEQ.ID.NO.:4 except that the nucleotide at position 120 is G, C, or T rather than A, so that the codon at positions 120-122 does not encode threonine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that at least one of the nucleotides at position 120 or
15 121 has been changed so that the codon at positions 120-122 does not encode threonine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 8,939 is A rather than T. Also included in the
20 present invention is a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 8,939 of SEQ.ID.NO.:1 is A, G, or C, rather than T, so that the codon at positions 8,939-8,941 does not encode tyrosine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is
25 identical to SEQ.ID.NO.:1 except that at least one of the nucleotides at position 8,939-8,941 has been changed so that the codon at positions 8,939-8,941 does not encode tyrosine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of
30 SEQ.ID.NO.:2 except that the nucleotide at position 783 is A rather than T. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that the nucleotide at position 783 is A, G, or C rather than T so that the codon at positions 783-785 does not encode
35 tyrosine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that at least one of the nucleotides at position 783-

785 has been changed so that the codon at positions 783-785 does not encode tyrosine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that the nucleotide at position 783 is A rather than T. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that the nucleotide at position 783 is A, G, or C rather than T, so that the codon at positions 783-785 does not encode tyrosine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,409 of SEQ.ID.NO.:4 except that at least one of the nucleotides at position 783-785 has been changed so that the codon at positions 783-785 does not encode tyrosine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 11,241 is A rather than G. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that the nucleotide at position 11,241 is A, C, or T, rather than G, so that the codon at positions 11,240-11,242 does not encode glycine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to SEQ.ID.NO.:1 except that at least one of the nucleotides at position 11,240 or 11,241 has been changed so that the codon at positions 11,240-11,242 does not encode glycine.

The present invention includes a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that the nucleotide at position 1,000 is A rather than G. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that the nucleotide at position 1,000 is A, C, or T rather than G, so that the codon at positions 999-1,001 does not encode glycine. Also included in the present invention is a DNA molecule having a nucleotide sequence that is identical to positions 105-1,859 of SEQ.ID.NO.:2 except that at least one of the nucleotides at position 999 or 1,000 has been changed so that the codon at positions 999-1,001 does not

encode glycine. Another aspect of the present invention includes host cells that have been engineered to contain and/or express DNA sequences encoding CG1CE protein. Such recombinant host cells can be cultured under suitable conditions to produce CG1CE protein. An expression vector containing DNA encoding CG1CE protein can be used for expression of CG1CE protein in a recombinant host cell.

Recombinant host cells may be prokaryotic or eukaryotic, including but not limited to, bacteria such as *E. coli*, fungal cells such as yeast, mammalian cells including, but not limited to, cell lines of human, bovine, porcine, monkey and rodent origin, and insect cells including but not limited to *Drosophila* and silkworm derived cell lines. Cell lines derived from mammalian species which are suitable for recombinant expression of CG1CE protein and which are commercially available, include but are not limited to, L cells L-M(TK⁻) (ATCC CCL 1.3), L cells L-M (ATCC CCL 1.2), 293 (ATCC CRL 1573), Raji (ATCC CCL 86), CV-1 (ATCC CCL 70), COS-1 (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C127I (ATCC CRL 1616), BS-C-1 (ATCC CCL 26) and MRC-5 (ATCC CCL 171).

A variety of mammalian expression vectors can be used to express recombinant CG1CE in mammalian cells. Commercially available mammalian expression vectors which are suitable include, but are not limited to, pMC1neo (Stratagene), pSG5 (Stratagene), pcDNAI and pcDNAIamp, pcDNA3, pcDNA3.1, pCR3.1 (Invitrogen), EBO-pSV2-neo (ATCC 37593), pBPV-1(8-2) (ATCC 37110), pdBPV-MMTneo(342-12) (ATCC 37224), pRSVgpt (ATCC 37199), pRSVneo (ATCC 37198), and pSV2-dhfr (ATCC 37146). Following expression in recombinant cells, CG1CE can be purified by conventional techniques to a level that is substantially free from other proteins.

The present invention includes CG1CE protein substantially free from other proteins. The amino acid sequence of the full-length CG1CE protein is shown in Figure 3 as SEQ.ID.NO.:3. Thus, the present invention includes CG1CE protein substantially free from other proteins having the amino acid sequence SEQ.ID.NO.:3. Also included in the present invention is a CG1CE protein that is produced

from an alternatively spliced CG1CE mRNA where the protein has the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5.

5 Mutated forms of CG1CE proteins are intended to be within the scope of the present invention. In particular, mutated forms of SEQ.ID.NOs.:3 and 5 that give rise to Best's macular dystrophy are within the scope of the present invention. Accordingly, the present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 93 is cysteine rather than tryptophan. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 93 is cysteine rather than tryptophan. The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 93 is not tryptophan. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 93 is not tryptophan.

10 The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 85 is histidine rather than tyrosine. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 85 is histidine rather than tyrosine. The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 85 is not tyrosine. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 85 is not tyrosine.

20 The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 6 is proline rather than threonine. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 6 is proline rather than threonine. The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 6 is not threonine. The present invention also includes a protein having the amino acid sequence shown

in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 6 is not threonine.

5 The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 227 is asparagine rather than tyrosine. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 227 is asparagine rather than tyrosine. The present invention includes a protein having the amino acid sequence shown in Figure 3 as
10 SEQ.ID.NO.:3 except that the amino acid at position 227 is not tyrosine. The present invention also includes a protein having the amino acid sequence shown in Figure 5 as SEQ.ID.NO.:5 except that the amino acid at position 227 is not tyrosine.

15 The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 299 is glutamate rather than glycine. The present invention includes a protein having the amino acid sequence shown in Figure 3 as SEQ.ID.NO.:3 except that the amino acid at position 299 is not glycine. As with many proteins, it is possible to modify many of the
20 amino acids of CG1CE and still retain substantially the same biological activity as the original protein. Thus, the present invention includes modified CG1CE proteins which have amino acid deletions, additions, or substitutions but that still retain substantially the same biological activity as CG1CE. It is generally accepted that single amino acid
25 substitutions do not usually alter the biological activity of a protein (see, *e.g.*, Molecular Biology of the Gene, Watson *et al.*, 1987, Fourth Ed., The Benjamin/Cummings Publishing Co., Inc., page 226; and Cunningham & Wells, 1989, Science 244:1081-1085). Accordingly, the present invention includes polypeptides where one amino acid substitution has been made
30 in SEQ.ID.NOs.:3 or 5 wherein the polypeptides still retain substantially the same biological activity as CG1CE. The present invention also includes polypeptides where two amino acid substitutions have been made in SEQ.ID.NOs.:3 or 5 wherein the polypeptides still retain substantially the same biological activity as CG1CE. In particular, the
35 present invention includes embodiments where the above-described substitutions are conservative substitutions. In particular, the present

invention includes embodiments where the above-described substitutions do not occur in positions where the amino acid present in CG1CE is also present in one of the *C. elegans* proteins whose partial sequence is shown in Figure 7.

5 The CG1CE proteins of the present invention may contain post-translational modifications, *e.g.*, covalently linked carbohydrate.

 The present invention also includes chimeric CG1CE proteins. Chimeric CG1CE proteins consist of a contiguous polypeptide sequence of at least a portion of a CG1CE protein fused to a polypeptide
10 sequence of a non- CG1CE protein.

 The present invention also includes isolated forms of CG1CE proteins and CG1CE DNA. By "isolated CG1CE protein" or "isolated CG1CE DNA" is meant CG1CE protein or DNA encoding CG1CE protein that has been isolated from a natural source. Use of the
15 term "isolated" indicates that CG1CE protein or CG1CE DNA has been removed from its normal cellular environment. Thus, an isolated CG1CE protein may be in a cell-free solution or placed in a different cellular environment from that in which it occurs naturally. The term isolated does not imply that an isolated CG1CE protein is the only protein
20 present, but instead means that an isolated CG1CE protein is at least 95% free of non-amino acid material (*e.g.*, nucleic acids, lipids, carbohydrates) naturally associated with the CG1CE protein. Thus, a CG1CE protein that is expressed in bacteria or even in eukaryotic cells which do not naturally (*i.e.*, without human intervention) express it
25 through recombinant means is an "isolated CG1CE protein."

 A cDNA fragment encoding full-length CG1CE can be isolated from a human retinal cell cDNA library by using the polymerase chain reaction (PCR) employing suitable primer pairs. Such primer pairs can be selected based upon the cDNA sequence for
30 CG1CE shown in Figure 2 as SEQ.ID.NO.:2 or in Figure 4 as SEQ.ID.NO.:4. Suitable primer pairs would be, *e.g.*:

 CAGGGAGTCCCACCAGCC (SEQ.ID.NO.:6) and

 TCCCCATTAGGAAGCAGG (SEQ.ID.NO.:7)

 for SEQ.ID.NO.:2; and

35 CAGGGAGTCCCACCAGCC (SEQ.ID.NO.:6) and

 TCTCCTCTTTGTTTCAGGC (SEQ.ID.NO.:8)

for SEQ.ID.NO.:4.

PCR reactions can be carried out with a variety of thermostable enzymes including but not limited to AmpliTaq, AmpliTaq Gold, or Vent polymerase. For AmpliTaq, reactions can be carried out in 10 mM Tris-Cl, pH 8.3, 2.0 mM MgCl₂, 200 μ M for each dNTP, 50 mM KCl, 0.2 μ M for each primer, 10 ng of DNA template, 0.05 units/ μ l of AmpliTaq. The reactions are heated at 95°C for 3 minutes and then cycled 35 times using the cycling parameters of 95°C, 20 seconds, 62°C, 20 seconds, 72°C, 3 minutes. In addition to these conditions, a variety of suitable PCR protocols can be found in PCR Primer, A Laboratory Manual, edited by C.W. Dieffenbach and G.S. Dveksler, 1995, Cold Spring Harbor Laboratory Press; or PCR Protocols: A Guide to Methods and Applications, Michael *et al.*, eds., 1990, Academic Press .

A suitable cDNA library from which a clone encoding CG1CE can be isolated would be Human Retina 5'-stretch cDNA library in lambda gt10 or lambda gt11 vectors (catalog numbers HL1143a and HL1132b, Clontech, Palo Alto, CA). The primary clones of such a library can be subdivided into pools with each pool containing approximately 20,000 clones and each pool can be amplified separately.

By this method, a cDNA fragment encoding an open reading frame of 585 amino acids (SEQ.ID.NO.:3) or an open reading frame of 435 amino acids (SEQ.ID.NO.:5) can be obtained. This cDNA fragment can be cloned into a suitable cloning vector or expression vector. For example, the fragment can be cloned into the mammalian expression vector pcDNA3.1 (Invitrogen, San Diego, Ca). CG1CE protein can then be produced by transferring an expression vector encoding CG1CE or portions thereof into a suitable host cell and growing the host cell under appropriate conditions. CG1CE protein can then be isolated by methods well known in the art.

As an alternative to the above-described PCR method, a cDNA clone encoding CG1CE can be isolated from a cDNA library using as a probe oligonucleotides specific for CG1CE and methods well known in the art for screening cDNA libraries with oligonucleotide probes. Such methods are described in, *e.g.*, Sambrook *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, *DNA Cloning: A*

Practical Approach, MRL Press, Ltd., Oxford, U.K., Vol. I, II.

Oligonucleotides that are specific for CG1CE and that can be used to screen cDNA libraries can be readily designed based upon the cDNA sequence of CG1CE shown in Figure 2 as SEQ.ID.NO.:2 or in Figure 4 as SEQ.ID.NO.:4 and can be synthesized by methods well-known in the art.

Genomic clones containing the CG1CE gene can be obtained from commercially available human PAC or BAC libraries available from Research Genetics, Huntsville, AL. PAC clones containing the CG1CE gene (e.g., PAC 759J12, PAC 466A11) are commercially available from Research Genetics, Huntsville, AL (Catalog number for individual PAC clones is RPCI.C). Alternatively, one may prepare genomic libraries, especially in P1 artificial chromosome vectors, from which genomic clones containing the CG1CE can be isolated, using probes based upon the CG1CE sequences disclosed herein. Methods of preparing such libraries are known in the art (Ioannou *et al.*, 1994, *Nature Genet.* 6:84-89).

The novel DNA sequences of the present invention can be used in various diagnostic methods relating to Best's macular dystrophy. The present invention provides diagnostic methods for determining whether a patient carries a mutation in the CG1CE gene that predisposes that patient toward the development of Best's macular dystrophy. In broad terms, such methods comprise determining the DNA sequence of a region of the CG1CE gene from the patient and comparing that sequence to the sequence from the corresponding region of the CG1CE gene from a normal person, *i.e.*, a person who does not suffer from Best's macular dystrophy.

Such methods of diagnosis may be carried out in a variety of ways. For example, one embodiment comprises:

(a) providing PCR primers from a region of the CG1CE gene where it is suspected that a patient harbors a mutation in the CG1CE gene;

(b) performing PCR on a DNA sample from the patient to produce a PCR fragment from the patient;

(c) performing PCR on a control DNA sample having a nucleotide sequence selected from the group consisting of SEQ.ID.NOs.:1, 2 and SEQ.ID.NO.:4 to produce a control PCR fragment;

(d) determining the nucleotide sequence of the PCR fragment from the patient and the nucleotide sequence of the control PCR fragment;

5 (e) comparing the nucleotide sequence of the PCR fragment from the patient to the nucleotide sequence of the control PCR fragment;

where a difference between the nucleotide sequence of the PCR fragment from the patient and the nucleotide sequence of the control PCR fragment indicates that the patient has a mutation in the
10 CG1CE gene.

In a particular embodiment, the PCR primers are from the coding region of the CG1CE gene, *i.e.*, from the coding region of SEQ.ID.NO.s.:1, 2, or 4.

15 In a particular embodiment, the DNA sample from the patient is cDNA that has been prepared from an RNA sample from the patient. In another embodiment, the DNA sample from the patient is genomic DNA.

In a particular embodiment, the nucleotide sequences of the PCR fragment from the patient and the control PCR fragment are
20 determined by DNA sequencing.

In a particular embodiment, the nucleotide sequences of the PCR fragment from the patient and the control PCR fragment are compared by direct comparison after DNA sequencing. In another embodiment, the comparison is made by a process that includes
25 hybridizing the PCR fragment from the patient and the control PCR fragment and then using an endonuclease that cleaves at any mismatched positions in the hybrid but does not cleave the hybrid if the two fragments match perfectly. Such an endonuclease is, *e.g.*, S1. In this embodiment, the conversion of the PCR fragment from the patient to
30 smaller fragments after endonuclease treatment indicates that the patient carries a mutation in the CG1CE gene. In such embodiments, it may be advantageous to label (radioactively, enzymatically, immunologically, *etc.*) the PCR fragment from the patient or the control PCR fragment.

35 The present invention provides a method of diagnosing whether a patient carries a mutation in the CG1CE gene that comprises:

(a) obtaining an RNA sample from the patient;
(b) performing reverse transcription-PCR (RT-PCR) on the RNA sample using primers that span a region of the coding sequence of the CG1CE gene to produce a PCR fragment from the patient
5 where the PCR fragment from the patient has a defined length, the length being dependent upon the identity of the primers that were used in the RT-PCR;

(c) hybridizing the PCR fragment to DNA having a sequence selected from the group consisting of SEQ.ID.NO.:1, 2 and
10 SEQ.ID.NO.:4 to form a hybrid ;

(d) treating the hybrid produced in step (c) with an endonuclease that cleaves at any mismatched positions in the hybrid but does not cleave the hybrid if the two fragments match perfectly;

(e) determining whether the endonuclease cleaved the
15 hybrid by determining the length of the PCR fragment from the patient after endonuclease treatment where a reduction in the length of the PCR fragment from the patient after endonuclease treatment indicates that the patient carries a mutation in the CG1CE gene.

The present invention provides a method of diagnosing
20 whether a patient carries a mutation in the CG1CE gene that comprises:

(a) making cDNA from an RNA sample from the patient;

(b) providing a set of PCR primers based upon
SEQ.ID.NO.:2 or SEQ.ID.NO.:4;

(c) performing PCR on the cDNA to produce a PCR
25 fragment from the patient;

(d) determining the nucleotide sequence of the PCR fragment from the patient;

(e) comparing the nucleotide sequence of the PCR
30 fragment from the patient with the nucleotide sequence of SEQ.ID.NO.:2 or SEQ.ID.NO.:4;

where a difference between the nucleotide sequence of the PCR fragment from the patient with the nucleotide sequence of
SEQ.ID.NO.:2 or SEQ.ID.NO.:4 indicates that the patient carries a
35 mutation in the CG1CE gene.

The present invention provides a method of diagnosing whether a patient carries a mutation in the CG1CE gene that comprises:

- (a) preparing genomic DNA from the patient;
- (b) providing a set of PCR primers based upon
5 SEQ.ID.NO.:1, SEQ.ID.NO.:2, or SEQ.ID.NO.:4;
- (c) performing PCR on the genomic DNA to produce a PCR fragment from the patient;
- (d) determining the nucleotide sequence of the PCR fragment from the patient;
- 10 (e) comparing the nucleotide sequence of the PCR fragment from the patient with the nucleotide sequence of SEQ.ID.NO.:2 or SEQ.ID.NO.:4;

where a difference between the nucleotide sequence of the PCR fragment from the patient with the nucleotide sequence of
15 SEQ.ID.NO.:2 or SEQ.ID.NO.:4 indicates that the patient carries a mutation in the CG1CE gene.

In a particular embodiment, the primers are selected so that they amplify a portion of SEQ.ID.NOs.:2 or 4 that includes at least one position selected from the group consisting of: positions 120, 121, 122,
20 357, 358, 359, 381, 382, 383, 783, 784, and 785. In another embodiment, the primers are selected so that they amplify a portion of SEQ.ID.NOs.:2 or 4 that includes at least one position selected from the group consisting of: positions 384, 385, and 386. In another embodiment, the primers are selected so that they amplify a portion of SEQ.ID.NO.:2 that includes at
25 least one position selected from the group consisting of: positions 999, 1,000, and 1,001. In another embodiment, the primers are selected so that they amplify a portion of SEQ.ID.NOs.:2 or 4 that includes at least one codon that encodes an amino acid present in CG1CE that is also present in the corresponding position in at least one of the *C. elegans*
30 proteins whose partial amino acid sequence is shown in Figure 7.

In a particular embodiment, the present invention provides a diagnostic method for determining whether a person carries a mutation of the CG1CE gene in which the G at position 383 of
SEQ.ID.NO.:2 has been changed to a C. This change results in the
35 creation of a Fnu4HI restriction site. By amplifying a PCR fragment spanning position 383 of SEQ.ID.NO.:2 from DNA or cDNA prepared

from a person, digesting the PCR fragment with Fnu4HI, and visualizing the digestion products, *e.g.*, by SDS-PAGE, one can easily determine if the person carries the G383C mutation. For example, one could use the PCR primer pair 5'-CTCCTGCCCAGGCTTCTAC-3' (SEQ.ID.NO.:30) and 5'-CTTGCTCTGCCTTGCCTTC-3' (SEQ.ID.NO.:31) to amplify a 125 base pair fragment. Heterozygotes for the G383C mutation have three Fnu4HI digestion products: 125 bp, 85 bp, and 40 bp; homozygotes have two: 85 bp and 40 bp; and wild-type individuals have a single fragment of 125 bp.

In a particular embodiment, the present invention provides a diagnostic method for determining whether a person carries a mutation of the CG1CE gene in which the T at position 783 of SEQ.ID.NO.:2 has been changed to an A. This change results in the creation of a PflMI restriction site. By amplifying a PCR fragment spanning position 783 of SEQ.ID.NO.:2 from DNA or cDNA prepared from a person, digesting the PCR fragment with PflMI, and visualizing the digestion products, *e.g.*, by SDS-PAGE, one can easily determine if the person carries the T783A mutation.

The present invention also provides oligonucleotide probes, based upon the sequences of SEQ.ID.NOs.:1, 2, or 4, that can be used in diagnostic methods related to Best's macular dystrophy. In particular, the present invention includes DNA oligonucleotides comprising at least 18 contiguous nucleotides of at least one of a sequence selected from the group consisting of: SEQ.ID.NOs.:1, 2 and SEQ.ID.:NO.4. Also provided by the present invention are corresponding RNA oligonucleotides. The DNA or RNA oligonucleotide probes can be packaged in kits.

In addition to the diagnostic utilities described above, the present invention makes possible the recombinant expression of the CG1CE protein in various cell types. Such recombinant expression makes possible the study of this protein so that its biochemical activity and its role in Best's macular dystrophy can be elucidated.

The present invention also makes possible the development of assays which measure the biological activity of the CG1CE protein. Such assays using recombinantly expressed CG1CE protein are especially of interest. Assays for CG1CE protein activity can be used to screen libraries of compounds or other sources of compounds to identify

compounds that are activators or inhibitors of the activity of CG1CE protein. Such identified compounds can serve as "leads" for the development of pharmaceuticals that can be used to treat patients having Best's macular dystrophy. In versions of the above-described
5 assays, mutant CG1CE proteins are used and inhibitors or activators of the activity of the mutant CG1CE proteins are discovered.

Such assays comprise:

- (a) recombinantly expressing CG1CE protein or mutant CG1CE protein in a host cell;
- 10 (b) measuring the biological activity of CG1CE protein or mutant CG1CE protein in the presence and in the absence of a substance suspected of being an activator or an inhibitor of CG1CE protein or mutant CG1CE protein;

where a change in the biological activity of the CG1CE
15 protein or the mutant CG1CE protein in the presence as compared to the absence of the substance indicates that the substance is an activator or an inhibitor of CG1CE protein or mutant CG1CE protein.

The present invention also includes antibodies to the CG1CE protein. Such antibodies may be polyclonal antibodies or
20 monoclonal antibodies. The antibodies of the present invention are raised against the entire CG1CE protein or against suitable antigenic fragments of the protein that are coupled to suitable carriers, *e.g.*, serum albumin or keyhole limpet hemocyanin, by methods well known in the art. Methods of identifying suitable antigenic fragments of a
25 protein are known in the art. See, *e.g.*, Hopp & Woods, 1981, Proc. Natl. Acad. Sci. USA 78:3824-3828; and Jameson & Wolf, 1988, CABIOS (Computer Applications in the Biosciences) 4:181-186.

For the production of polyclonal antibodies, CG1CE protein or an antigenic fragment, coupled to a suitable carrier, is injected on a
30 periodic basis into an appropriate non-human host animal such as, *e.g.*, rabbits, sheep, goats, rats, mice. The animals are bled periodically and sera obtained are tested for the presence of antibodies to the injected antigen. The injections can be intramuscular, intraperitoneal, subcutaneous, and the like, and can be accompanied with adjuvant.

35 For the production of monoclonal antibodies, CG1CE protein or an antigenic fragment, coupled to a suitable carrier, is

injected into an appropriate non-human host animal as above for the production of polyclonal antibodies. In the case of monoclonal antibodies, the animal is generally a mouse. The animal's spleen cells are then immortalized, often by fusion with a myeloma cell, as described
5 in Kohler & Milstein, 1975, Nature 256:495-497. For a fuller description of the production of monoclonal antibodies, see Antibodies: A Laboratory Manual, Harlow & Lane, eds., Cold Spring Harbor Laboratory Press, 1988.

Gene therapy may be used to introduce CG1CE polypeptides
10 into the cells of target organs, *e.g.*, the pigmented epithelium of the retina or other parts of the retina. Nucleotides encoding CG1CE polypeptides can be ligated into viral vectors which mediate transfer of the nucleotides by infection of recipient cells. Suitable viral vectors include retrovirus, adenovirus, adeno-associated virus, herpes virus,
15 vaccinia virus, and polio virus based vectors. Alternatively, nucleotides encoding CG1CE polypeptides can be transferred into cells for gene therapy by non-viral techniques including receptor-mediated targeted transfer using ligand-nucleotide conjugates, lipofection, membrane fusion, or direct microinjection. These procedures and variations
20 thereof are suitable for *ex vivo* as well as *in vivo* gene therapy. Gene therapy with CG1CE polypeptides will be particularly useful for the treatment of diseases where it is beneficial to elevate CG1CE activity.

The present invention includes DNA comprising nucleotides encoding mouse CG1CE. Included within such DNA is the
25 DNA sequence shown in Figure 8A-C (SEQ. ID. NO.:28). Also included is DNA comprising positions 11-1,663 of SEQ. ID. NO.:28. Also included are mutant versions of DNA encoding mouse CG1CE. Included is DNA comprising nucleotides that are identical to positions 11-1,663 of SEQ. ID. NO.:28 except that at least one of the nucleotides at positions 26-28,
30 positions 263-265, positions 287-289, positions 689-691, and/or positions 905-907 differs from the corresponding nucleotide at positions 26-28, positions 263-265, positions 287-289, positions 689-691, and/or positions 905-907 of SEQ. ID. NO.:28. Particularly preferred versions of mutant DNAs are those in which the nucleotide change results in a change in
35 the corresponding encoded amino acid. The DNA encoding mouse

CG1CE can be in isolated form, can be substantially free from other nucleic acids, and/or can be recombinant DNA.

The present invention includes mouse CG1CE protein (SEQ. ID. NO.:29). This mouse CG1CE protein can be in isolated form and/or can be substantially free from other proteins. Mutant versions of mouse CG1CE protein are also part of the present invention. Examples of such mutant mouse CG1CE proteins are proteins that are identical to SEQ. ID. NO.:29 except that the amino acid at position 6, position 85, position 93, position 227, and/or position 299 differs from the corresponding amino acid at position 6, position 85, position 93, position 227, and/or position 299 in SEQ. ID. NO.:29.

cDNA encoding mouse CG1CE can be amplified by PCR from cDNA libraries made from mouse eye or mouse testis. Suitable primers can be readily designed based upon SEQ. ID. NO.:28. Alternatively, cDNA encoding mouse CG1CE can be isolated from cDNA libraries made from mouse eye or mouse testis by the use of oligonucleotide probes based upon SEQ. ID. NO.:28.

In situ hybridization studies demonstrated that mouse CG1CE is specifically expressed in the retinal pigmented epithelium (see Figure 10).

By providing DNA encoding mouse CG1CE, the present invention allows for the generation of an animal model of Best's macular dystrophy. This animal model can be generated by making "knockout" or "knockin" mice containing altered CG1CE genes. Knockout mice can be generated in which portions of the mouse CG1CE gene have been deleted. Knockin mice can be generated in which mutations that have been shown to lead to Best's macular dystrophy when present in the human CG1CE gene are introduced into the mouse gene. In particular, mutations resulting in changes in amino acids 6, 85, 93, 227, or 299 of the mouse CG1CE protein (SEQ.ID.NO.:29) are contemplated. Such knockout and knockin mice will be valuable tools in the study of the Best's macular dystrophy disease process and will provide important model systems in which to test potential pharmaceuticals or treatments for Best's macular dystrophy.

Methods of producing knockout and knockin mice are well known in the art. For example, the use of gene-targeted ES cells

in the generation of gene-targeted transgenic knockout mice is described in, *e.g.*, Thomas et al., 1987, Cell 51:503-512, and is reviewed elsewhere (Frohman et al., 1989, Cell 56:145-147; Capecchi, 1989, Trends in Genet. 5:70-76; Baribault et al., 1989, Mol. Biol. Med. 5 6:481-492).

Techniques are available to inactivate or alter any genetic region to virtually any mutation desired by using targeted homologous recombination to insert specific changes into chromosomal genes. Generally, use is made of a "targeting vector,"
10 *i.e.*, a plasmid containing part of the genetic region it is desired to mutate. By virtue of the homology between this part of the genetic region on the plasmid and the corresponding genetic region on the chromosome, homologous recombination can be used to insert the plasmid into the genetic region, thus disrupting the genetic region.
15 Usually, the targeting vector contains a selectable marker gene as well.

In comparison with homologous extrachromosomal recombination, which occurs at frequencies approaching 100%, homologous plasmid-chromosome recombination was originally
20 reported to only be detected at frequencies between 10^{-6} and 10^{-3} (Lin et al., 1985, Proc. Natl. Acad. Sci. USA 82:1391-1395; Smithies et al., 1985, Nature 317: 230-234; Thomas et al., 1986, Cell 44:419-428). Nonhomologous plasmid-chromosome interactions are more frequent, occurring at levels 10^5 -fold (Lin et al., 1985, Proc. Natl.
25 Acad. Sci. USA 82:1391-1395) to 10^2 -fold (Thomas et al., 1986, Cell 44:419-428) greater than comparable homologous insertion.

To overcome this low proportion of targeted recombination in murine ES cells, various strategies have been developed to detect or select rare homologous recombinants. One
30 approach for detecting homologous alteration events uses the polymerase chain reaction (PCR) to screen pools of transformant cells for homologous insertion, followed by screening individual clones (Kim et al., 1988, Nucleic Acids Res. 16:8887-8903; Kim et al., 1991, Gene 103:227-233). Alternatively, a positive genetic selection
35 approach has been developed in which a marker gene is constructed which will only be active if homologous insertion occurs, allowing

these recombinants to be selected directly (Sedivy et al., 1989, Proc. Natl. Acad. Sci. USA 86:227-231). One of the most powerful approaches developed for selecting homologous recombinants is the positive-negative selection (PNS) method developed for genes for which no direct selection of the alteration exists (Mansour et al., 1988, Nature 336:348-352; Capecchi, 1989, Science 244:1288-1292; Capecchi, 1989, Trends in Genet. 5:70-76). The PNS method is more efficient for targeting genes which are not expressed at high levels because the marker gene has its own promoter. Nonhomologous recombinants are selected against by using the Herpes Simplex virus thymidine kinase (HSV-TK) gene and selecting against its nonhomologous insertion with herpes drugs such as gancyclovir (GANC) or FIAU (1-(2-deoxy 2-fluoro-B-D-arabino)fluranosyl)-5-iodouracil). By this counter-selection, the percentage of homologous recombinants in the surviving transformants can be increased.

The following non-limiting examples are presented to better illustrate the invention.

20

EXAMPLE 1

Identification of the human CG1CE gene and cDNA cloning

Construction of Libraries for Shotgun Sequencing

Bacterial strains containing the BMD PACs (P1 Artificial Chromosomes) were received from Research Genetics (Huntsville, AL). The minimum tiling path between markers D11S4076 and UGB that represents the minimum genetic region containing the BMD gene includes the following nine PAC clones: 363M5 (140 kb), 519O13(120 kb), 527E4 (150 kb), 688P12 (140 kb), 741N15 (170 kb), 756B9 (120 kb), 759J12 (140 kb), 1079D9 (170 kb), and 363P2 (160 kb). Cells were streaked on Luria-Bertani (LB) agar plates supplemented with the appropriate antibiotic. A single colony was picked up and subjected to colony-PCR analysis with corresponding STS primers described in Cooper *et al.*, 1997, Genomics 41:185-192 to confirm the authenticity of PAC clones. A single positive colony was used to prepare a 5-ml starter culture and then 1-L overnight

culture in LB medium. The cells were pelleted by centrifugation and PAC DNA was purified by equilibrium centrifugation in cesium chloride-ethidium bromide gradient (Sambrook, Fritsch, and Maniatis, 1989, Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Harbor Laboratory Press). Purified PAC DNA was brought to 50 mM Tris pH 8.0, 15 mM MgCl₂, and 25% glycerol in a volume of 2 ml and placed in a AERO-MIST nebulizer (CIS-US, Bedford, MA). The nebulizer was attached to a nitrogen gas source and the DNA was randomly sheared at 10 psi for 30 sec. The sheared DNA was ethanol precipitated and resuspended in TE (10 mM Tris, 1 mM EDTA). The ends were made blunt by treatment with Mung Bean Nuclease (Promega, Madison, WI) at 30°C for 30 min, followed by phenol/chloroform extraction, and treatment with T4 DNA polymerase (GIBCO/BRL, Gaithersburg, MD) in multicore buffer (Promega, Madison, WI) in the presence of 40 uM dNTPs at 16°C. To facilitate subcloning of the DNA fragments, BstX I adapters (Invitrogen, Carlsbad, CA) were ligated to the fragments at 14°C overnight with T4 DNA ligase (Promega, Madison, WI). Adapters and DNA fragments less than 500 bp were removed by column chromatography using a cDNA sizing column (GIBCO/BRL, Gaithersburg, MD) according to the instructions provided by the manufacturer. Fractions containing DNA greater than 1 kb were pooled and concentrated by ethanol precipitation. The DNA fragments containing BstX I adapters were ligated into the BstX I sites of pSHOT II which was constructed by subcloning the BstX I sites from pcDNA II (Invitrogen, Carlsbad, CA) into the BssH II sites of pBlueScript (Stratagene, La Jolla, CA). pSHOT II was prepared by digestion with BstX I restriction endonuclease and purified by agarose gel electrophoresis. The gel purified vector DNA was extracted from the agarose by following the Prep-A-Gene (BioRad, Richmond, CA) protocol. To reduce ligation of the vector to itself, the digested vector was treated with calf intestinal phosphatase (GIBCO/BRL, Gaithersburg, MD). Ligation reactions of the DNA fragments with the cloning vector were transformed into ultra-competent XL-2 Blue cells (Stratagene, La Jolla, CA), and plated on LB agar plates supplemented with 100 µg/ml ampicillin. Individual colonies were picked into a 96 well plate containing 100 µl/well of LB broth supplemented with ampicillin and

grown overnight at 37°C. Approximately 25 µl of 80% sterile glycerol was added to each well and the cultures stored at -80°C.

Preparation of plasmid DNA

5 Glycerol stocks were used to inoculate 5 ml of LB broth supplemented with 100 µg/ml ampicillin either manually or by using a Tecan Genesis RSP 150 robot (Tecan AG, Hombrechtikon, Switzerland) programmed to inoculate 96 tubes containing 5 ml broth from the 96 wells. The cultures were grown overnight at 37°C with shaking to
10 provide aeration. Bacterial cells were pelleted by centrifugation, the supernatant decanted, and the cell pellet stored at -20°C. Plasmid DNA was prepared with a QIAGEN Bio Robot 9600 (QIAGEN, Chatsworth, CA) according to the Qiawell Ultra protocol. To test the frequency and size of inserts, plasmid DNA was digested with the restriction
15 endonuclease Pvu II. The size of the restriction endonuclease products was examined by agarose gel electrophoresis with the average insert size being 1 to 2 kb.

DNA Sequence Analysis of Shotgun clones

20 DNA sequence analysis was performed using the ABI PRISM™ dye terminator cycle sequencing ready reaction kit with AmpliTaq DNA polymerase, FS (Perkin Elmer, Norwalk, CT). DNA sequence analysis was performed with M13 forward and reverse primers. Following amplification in a Perkin-Elmer 9600, the extension
25 products were purified and analyzed on an ABI PRISM 377 automated sequencer (Perkin Elmer, Norwalk, CT). Approximately 4 sequencing reactions were performed per kb of DNA to be examined (384 sequencing reactions per each of nine PACs).

Assembly of DNA sequences

30 Phred/Phrap was used for DNA sequences assembly. This program was developed by Dr. Phil Green and licensed from the University of Washington (Seattle, WA). Phred/Phrap consists of the following programs: Phred for base-calling, Phrap for sequence
35 assembly, Crossmatch for sequence comparisons, Consed and

Phrapview for visualization of data, Repeatmasker for screening repetitive sequences. Vector and *E. coli* DNA sequences were identified by Crossmatch and removed from the DNA sequence assembly process. DNA sequence assembly was on a SUN Enterprise 4000 server running
5 a Solaris 2.51 operating system (Sun Microsystems Inc., Mountain View, CA) using default Phrap parameters. The sequence assemblies were further analyzed using Consed and Phrapview.

10 Identification of new microsatellite genetic markers from the Best's macular dystrophy region

Isolation of CA microsatellites from PAC-specific sublibraries, Southern blotting and hybridization of PAC DNA with a (dC-dA)_n-(dG-dT)_n probe (Pharmacia Biotech, Uppsala, Sweden) was used to confirm the presence of CA repeats in nine PAC clones that
15 represent a minimum tiling path. Shotgun PAC-specific sublibraries were constructed from DNA of all 9 PAC clones using a protocol described above. The sublibraries were plated on agar plates, and colonies were transferred to nylon membranes and probed with randomly primed polynucleotide, (dC-dA)_n-(dG-dT)_n. Hybridization was
20 performed overnight in a solution containing 6X SSC, 20 mM sodium phosphate buffer (pH 7.0), 1% bovine serum albumin, and 0.2% sodium dodecyl sulfate at 65°C. Filters were washed four times for 15 min each in 2X SSC and 0.2% SDS at 65°C. CA-positive subclones were identified for all but one PAC clone (527E4). DNA from these subclones was
25 isolated and sequenced as described above for the shotgun library clones.

Identification of simple repeat sequences in assembled DNA sequences. DNA sequence at the final stage of assembly was checked for the presence of microsatellite repeats using a Consed
30 visualization tool of the Phred/Phrap package.

Polymorphism analysis and recombination mapping

Sequence fragments containing CA repeats were analyzed using the PRIMER program; oligonucleotide pairs flanking each of the
35 CA repeats were synthesized. The forward primer was kinase-labeled with [gamma-32P]-ATP. Amplification of the genomic DNA was

performed in a total volume of 10 μ l containing 5 ng/ μ l of genomic DNA; 10 mM Tris-HCl pH 8.3; 1.5 mM MgCl₂; 50 mM KCl; 0.01% gelatin; 200 μ M dNTPs; 0.2 pmol/ μ l of both primers; 0.025 unit/ μ l of Taq polymerase. The PCR program consisted of 94° C for 3 min followed by 30 cycles of
5 94° C for 1 min, 55° C for 2 min, 72° C for 2 min and a final elongation step at 72° C for 10 min. Following amplification, samples were mixed with 2 vol of a formamide dye solution and run on a 6% polyacrylamide sequencing gel. Two newly identified markers detected two
10 recombination events in disease chromosomes of individuals from family S1. This limited the minimum genetic region to the interval covered by 6 PAC clones: 519O13, 759J12, 756B9, 363M5, 363P2, and 741N15.

Identification of the retina-specific EST hit in the pCA759J12-2 clone.

15 A CA-positive subclone (pCA759J12-2) was identified in the shotgun library generated from the PAC 759J12 DNA by hybridization to the (dC-dA)_n (dG-dT)_n probe. DNA sequence from pCA759J12-2 was queried against the EST sequences in the GenBank database using the BLAST algorithm (S.F. Altschul, *et al.*, 1990, J. Mol. Biol. 215:403-410).
20 The BLAST analysis identified a high degree of similarity between the DNA sequence obtained from the clone pCA759J12-2 and a retina-specific human EST with GenBank accession number AA318352. BLASTX analysis of EST AA318352 revealed a strong homology of the corresponding protein to a group of *C. elegans* proteins with unknown
25 function (RFP family). The RFP family is known only from *C. elegans* genome and EST sequences (e.g., *C. elegans* C29F4.2 and B0564.3) and is named for the amino acid sequence RFP that is invariant among 15 of the 16 family members; members share a conserved 300-400 amino acid sequence including 25 highly conserved aromatic residues.

30 A human gene partially represented in pCA759J12-2 and EST AA318352 was dubbed CG1CE (Candidate Gene #1 with the homology to the *C. elegans* group of genes) and selected for detailed analysis.

BioInformatic Analysis of Assembled DNA Sequences

When the assembled DNA sequences from the nine BMD PACs approached 0.5-1-fold coverage, the DNA contigs were randomly concatenated, and prediction abilities of the program package AceDB were utilized to aid in gene identification.

In addition to the DNA sequence generated from the nine PACs mentioned above, Genbank database entries for PACs 466A11 and 363P2 (GenBank accession numbers AC003025 and AC003023, respectively) were analyzed with the use of the same AceDB package. PAC clones 466A11 and 363P2 represent parts of the PAC contig across the BMD region (Cooper *et al.*, 1997, Genomics 41:185-192); both clones map to the minimum genetic region containing the BMD gene that was determined by recombination breakpoint analysis in a 12-generation Swedish pedigree (Graff *et al.*, 1997, Hum. Genet. 101: 263-279). Database entries for PACs 466A11 and 363P2 represent unordered DNA pieces generated in Phase 1 High Throughput Genome Sequence Project (HTGS phase 1) by Genome Science and Technology Center, University of Texas Southwestern Medical Center at Dallas.

cDNA sequence and exon/intron organization of the CG1CE gene

Genomic DNA sequences from PACs 466A11 and 759J12 were compared with the CG1CE cDNA sequence from EST AA318352 using the program Crossmatch which allowed for a rapid and sensitive detection of the location of exons. The identification of intron/exon boundaries was then accomplished by manually comparing visualized genomic and cDNA sequences by using the AceDB package. This analysis allowed the identification of exons 8, 9, and 10 that are represented in EST AA318352. To increase the accuracy of the analysis, the DNA sequence of EST AA318352 was verified by comparison with genomic sequence obtained from pCA759J12-2, PAC 466A11, and shotgun PAC 759J12 subclones. The verified EST AA318352 sequence was reanalyzed by BLAST; two new EST's (accession numbers AA307119 and AA205892) were found to partially overlap with EST AA318352. They were assembled into a contig using the program Sequencher (Perkin Elmer, Norwalk, CT), and a consensus sequence derived from three ESTs (AA318352, AA307119, and AA205892) was re-analyzed by BLAST.

BLAST analysis identified a fourth EST belonging to this cluster (accession number AA317489); EST AA317489 was included in the consensus cDNA sequence. The consensus sequence derived from the four ESTs (AA318352, AA307119, AA205892, and AA317489) was compared with genomic sequences obtained from pCA759J12-2, PAC 466A11, and shotgun PAC 759J12 subclones using the programs Crossmatch and AceDB. This analysis verified the sequence and corrected sequencing errors that were found in AA318352, AA307119, AA205892, and AA317489. Comparison of cDNA and genomic sequences revealed a total of 7 exons. The order of the exons from 5' end to 3' end was 5'-ex4-ex5-ex6-ex8-ex9-ex10-ex11-3'. BLASTX analysis of the genomic segment located between exons 6 and 8 in PAC 466A11 revealed strong homology of the corresponding protein to a group of *C. elegans* proteins (RFP family). Since there were no EST hits in the GenBank EST database that covers this stretch of genomic sequence, this part of the CG1CE gene was called exH (Hypothetical ex 7). This finding changed the order of exons in the CG1CE gene to 5'-ex4-ex5-ex6-ex7-ex8-ex9-ex10-ex11-3'. The BLAST analysis of the DNA region located upstream of the exon 4 identified an additional human EST (AA326727) with a high degree of similarity to genomic sequence. Comparison of DNA and genomic sequences revealed the presence of two additional exons (ex1 and ex2) in the CG1CE gene. This finding changed the order of the exons in the CG1CE gene to 5'-ex1-ex2-ex4-ex5-ex6-ex7-ex8-ex9-ex10-ex11-3'. Bioinformatic analysis did not allow the prediction of boundaries between exons 2 and 4, exons 6 and 7, and exons 7 and 8. In addition, there was no overlap between ESTs represented in exons 1 and 2 from one side and exons 4, 5, 6, 7, 8, 9, 10, and 11 from another. There was the possibility of the presence of additional exons in the CG1CE gene that were not represented in the GenBank EST database.

Identification of an additional exon and determination of the exact exon/intron boundaries within the CG1CE gene.

To identify additional exon(s) within the CG1CE gene and verify the exonic composition of this gene, forward and reverse PCR primers from all known exons of the CG1CE gene were synthesized and used to PCR amplify CG1CE cDNA fragments from human retina

“Marathon-ready” cDNA (Clontech, Palo Alto, CA). In these RT-PCR experiments forward primer from ex1 (LF: CTAGTCGCCAGACCTTCTGTG) (SEQ.ID.NO.:9) was paired with a reverse primer from ex4 (GR: CTTGTAGACTGCGGTGCTGA) (SEQ.ID.NO.:10), forward primer from ex4 (GF: GAAAGCAAGGACGAGCAAAG) (SEQ.ID.NO.:11) was paired with a reverse primer from ex6 (ER: AATCCAGTCGTAGGCATACAGG) (SEQ.ID.NO.:12), forward primer from ex6 (EF: ACCTTGCGTACTCAGTGTGGA) (SEQ.ID.NO.:13) was paired with a reverse primer from ex8 (AR: TGTCGACAATCCAGTTGGTCT) (SEQ.ID.NO.:14), forward primer from ex8 (AF: CCCTTTGGAGAGGATGATGA) (SEQ.ID.NO.:15) was paired with a reverse primer from ex10 (CR: CTCTGGCATATCCGTCAGGT) (SEQ.ID.NO.:16), forward primer from ex10 (CF: CTTCAAGTCTGCCCCACTGT) (SEQ.ID.NO.:17) was paired with a reverse primer from ex11 (DR: GCATCCCCATTAGGAAGCAG) (SEQ.ID.NO.:18).

A 50 µl PCR reaction was performed using the Taq Gold DNA polymerase (Perkin Elmer, Norwalk, CT) in the reaction buffer supplied by the manufacturer with the addition of dNTPs, primers, and approximately 0.5 ng of human retina cDNA. PCR products were electrophoresed on a 2% agarose gel and DNA bands were excised, purified and subjected to sequence analysis with the same primers that were used for PCR amplification. The assembly of the DNA sequence results of these PCR products revealed that:

- (i) exons 1 and 2 from one side and exons 4, 5, 6, 7, 8, 9, 10, and 11 indeed represent fragments of the same gene
- (ii) an additional exon is present between exons 2 and 4 (named ex3)
- (iii) exon 7 (Hypothetical) predicted by the BLASTX analysis is present in the CG1CE cDNA fragment amplified by EF/AR primers.

Comparison of the DNA sequences obtained from RT-PCR fragments with genomic sequences obtained from pCA759J12-2, PAC 466A11, and shotgun PAC 759J12 subclones was performed using the

programs Crossmatch and AceDB. This analysis confirmed the presence of the exons originally found in five ESTs (AA318352, AA307119, AA205892, AA317489, and AA326727) and identified an additional exon (exon3) in the CG1CE gene. Exact sequence of
5 exon/intron boundaries within the CG1CE gene were determined for all of the exons. The splice signals in all introns conform to publish consensus sequences. The CG1CE gene appears to span at least 16 kb of genomic sequence. It contains a total of 11 exons.

10 Two splice donor sites for intron 7.

Two splicing variants of exon 7 were detected upon sequence analysis of RT-PCR products amplified from human retina cDNA with the primer pair EF/AR. Two variants utilize alternative splice donor sites separated from each other by 203 bp. Both splicing
15 sites conform to the published consensus sequence.

Identification of 5' and 3' ends of CG1CE cDNA

RACE is an established protocol for the analysis of cDNA ends. This procedure was performed using the Marathon RACE
20 template from human retina, purchased from Clontech (Palo Alto, CA). cDNA primers KR (CTAAGCGGGCATTAGCCACT) (SEQ.ID.NO.:19) and LR(TGGGGTTCCAGGTGGGTCCGAT) (SEQ.ID.NO.:20) in combination with a cDNA adaptor primer AP1 (CCATCCTAATACGACTCACTATAGGGC) (SEQ.ID.NO.:21) were
25 used in 5'RACE. cDNA primer DF (GGATGAAGCACATTCTAACCTGCTTC) (SEQ.ID.NO.:22) in combination with a cDNA adaptor primer AP1 (CCATCCTAATACGACTCACTATAGGGC) (SEQ.ID.NO.:21) was used in 3'RACE. Products obtained from these PCR amplifications were
30 analyzed on 2% agarose gels. Excised fragments from the gels were purified using Qiagen QIAquick spin columns and sequenced using ABI dye-terminator sequencing kits. The products were analyzed on ABI 377 sequencers according to standard protocols.

EXAMPLE 2

Best's macular dystrophy is associated with mutations in an evolutionarily conserved region of CG1CE

5 Genomic DNA from BMD patients from two Swedish pedigrees having Best's macular dystrophy (families S1 and SL76) was amplified by PCR using the following primer pair:

exG_left AAAGCTGGAGGAGCCGAG (SEQ.ID.NO.:23)

exG_right CTCCACCCATCTTCCGTTC (SEQ.ID.NO.:24)

10 This primer pair amplifies a genomic fragment that is 412 bp long and contains exon4 and adjacent intronic regions.

The patients were:

Family S1:

S1-3, a normal individual, *i.e.*, not having BMD; sister of S1-4

S1-4, an individual heterozygous for BMD; and

15 S1-5, an individual homozygous for BMD.

Patients S1-4 and S1-5 had the clinical symptoms of BMD, including morphological changes observable upon ophthalmologic examination.

Family SL76:

SL76-3, an individual heterozygous for BMD; mother of SL76-2

20 SL76-2, an individual heterozygous for BMD, son of SL-3.

PCR products produced using the primer sets mentioned above were amplified in 50 µl reactions consisting of Perkin-Elmer 10 x PCR Buffer, 200 mM dNTP's, 0.5 ul of Taq Gold (Perkin-Elmer Corp., Foster City, CA), 50 ng of patient DNA and 0.2 µM of forward and
25 reverse primers. Cycling conditions were as follows:

1. 94°C 10 min
2. 94°C 30 sec
3. 72°C 2 min (decrease this temperature by 1.1°C per cycle)
4. 72°C 2 min
- 30 5. Go to step 2 15 more times
6. 94°C 30 sec
7. 55°C 2 min
8. 72°C 2 min
9. Go to step 6 24 more times

10. 72°C 7 min

11. 4°C

Products obtained from this PCR amplification were analyzed on 2% agarose gels and excised fragments from the gels were purified using Qiagen QIAquick spin columns and sequenced using ABI dye-terminator sequencing kits. The products were analyzed on ABI 377 sequencers according to standard protocols.

The results are shown in Figure 6. Figure 6 shows a chromatogram from sequencing runs on the PCR fragments from patients S1-3, S1-4, and S1-5. The six readings represent sequencing of both strands of the PCR fragments from the patients. As can be seen from Figure 6, the two patients affected with BMD, patients S1-4 and S1-5, both carry a mutation at position 383 of SEQ.ID.NO.:2. Both copies of the CG1CE gene are mutated in homozygous affected S1-5, while heterozygous affected S1-4 contains both normal and mutated copies of the CG1CE gene. This mutation changes the codon that encodes the amino acid at position 93 of SEQ.ID.NO.:3 from TGG (encoding tryptophan) to TGC (encoding cysteine). Patient S1-3, a normal individual, has the wild-type sequence, TGG, at this codon. This disease mutation that changes this TGG codon to a TGC codon was not found upon sequencing of 50 normal unrelated individuals (100 chromosomes) of North American descent.

Both patients from family SL76 carry a mutation at position 357 of SEQ.ID.NO.:2. This mutation changes the codon that encodes the amino acid at position 85 of SEQ.ID.NO.:3 from TAC (encoding tyrosine) to CAC (encoding histidine). This disease mutation that changes this TAC codon to a CAC codon was not found upon sequencing of 50 normal unrelated individuals (100 chromosomes) of North American descent.

Amino acid positions 85 and 93 of the CG1CE protein are evolutionarily conserved. Figure 7 demonstrates that position 93 is occupied by tryptophan not only in the CG1CE protein, but also in 15 of 16 related *C. elegans* proteins. The lone *C. elegans* protein in which this residue is not tryptophan contains an isofunctional phenylalanine instead. Phenylalanine and tryptophan, both being hydrophobic, aromatic amino acids, are highly similar. Position 85 is occupied by

tyrosine and isofunctional phenylalanine in all 16 related *C. elegans* proteins. Phenylalanine and tyrosine, both being aromatic amino acids, are highly similar.

5

EXAMPLE 3

Expression of CG1CE

RT-PCR: RT-PCR experiments were performed on "quick-clone" human cDNA samples available from Clontech, Palo Alto, CA. cDNA samples from heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, and retina were amplified with primers AF (CCCTTTGGAGAGGATGATGA) (SEQ.ID.NO.:15) and CR (CTCTGGCATATCCGTCAGGT) (SEQ.ID.NO.:16) in the following PCR conditions:

1. 94°C 10 min
- 15 2. 94°C 30 sec
3. 72°C 2 min (decrease this temperature by 1.1°C per cycle)
4. 72°C 2 min
5. Go to step 2 15 more times
6. 94°C 30 sec
- 20 7. 55°C 2 min
8. 72°C 2 min
9. Go to step 6 19 more times
10. 72°C 7 min
11. 4°C
- 25 The CG1CE gene was found to be predominantly expressed in human retina and brain

Northern blot analysis: Northern blots containing poly(A+)-RNA from different human tissues were purchased from Clontech, Palo Alto, CA. Blot #1 contained human heart, brain placenta, lung, liver, skeletal muscle, kidney, and pancreas poly(A+)-RNA. Blot #2 contained stomach, thyroid, spinal cord, lymph node, trachea, adrenal gland, and bone marrow poly(A+)-RNA.

Primers CF (CTTCAAGTCTGCCCCACTGT) (SEQ.ID.NO.:17) and
exC_right (TAGGCTCAGAGCAAGGGAAG) (SEQ.ID.NO.:25) were
used to amplify a PCR product from total genomic DNA. This product
was purified on an agarose gel, and used as a probe in Northern blot
5 hybridization. The probe was labeled by random priming with the
Amersham Rediprime kit (Arlington Heights, IL) in the presence of 50-
100 μ Ci of 3000 Ci/mmol [alpha 32 P]dCTP (Dupont/NEN, Boston, MA).
Unincorporated nucleotides were removed with a ProbeQuant G-50 spin
column (Pharmacia/Biotech, Piscataway, NJ). The radiolabeled probe
10 at a concentration of greater than 1×10^6 cpm/ml in rapid hybridization
buffer (Clontech, Palo Alto, CA) was incubated overnight at 65°C. The
blots were washed by two 15 min incubations in 2X SSC, 0.1% SDS
(prepared from 20X SSC and 20 % SDS stock solutions, Fisher,
Pittsburgh, PA) at room temperature, followed by two 15 min
15 incubations in 1X SSC, 0.1% SDS at room temperature, and two 30 min
incubations in 0.1X SSC, 0.1% SDS at 60°C. Autoradiography of the blots
was done to visualize the bands that specifically hybridized to the
radiolabeled probe.

The probe hybridized to an mRNA transcript that is
20 uniquely expressed in brain and spinal cord.

Mouse probe for the murine ortholog of the GC1CE gene
was generated based on the sequence of an EST with GenBank accession
number AA497726. The 246 bp probe was amplified from mouse heart
cDNA (Clontech, Palo Alto, CA) using the primers mouseCG1CE_L
25 (ACACAACACATTCTGGGTGC) (SEQ.ID.NO.:26) and
mouseCG1CE_R (TTCAGAAACTGCTTCCCGAT) (SEQ.ID.NO.:27).
Due to an extremely low expression level of the CG1CE gene in mouse
heart, repetitive amplification steps were used to generate this probe.
The authenticity of this probe was verified by sequence analysis of the gel
30 purified DNA band. Northern blot containing poly(A+)-RNA from
several rat tissues (heart, brain, spleen, lung, liver, skeletal muscle,
kidney, testis) was purchased from Clontech, Palo Alto, CA. The probe
hybridized to an mRNA transcript that is expressed in testis only.

The present invention is not to be limited in scope by the
35 specific embodiments described herein. Indeed, various modifications
of the invention in addition to those described herein will become

apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

5 Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

WHAT IS CLAIMED:

1. An isolated DNA comprising nucleotides encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ.ID.NO.:3, SEQ.ID.NO.:5, and SEQ.ID.NO.:29.
5
2. The DNA of claim 1 comprising a nucleotide sequence selected from the group consisting of: SEQ.ID.NO.:1, SEQ.ID.NO.:2, SEQ.ID.NO.:4, SEQ.ID.NO.:28, positions 105-1,859 of SEQ.ID.NO.:2, positions 105-1,409 of SEQ.ID.NO.:4, and positions 11-1,663 of SEQ.ID.NO.:28.
10
3. An isolated DNA comprising a sequence that is identical to SEQ.ID.NO.:2 except that it contains a different nucleotide at a position selected from the group consisting of positions 120, 121, 122, 357, 358, 359, 381, 382, 383, 783, 784, 785, 999, 1000, and 1001.
15
4. An isolated DNA that hybridizes under stringent conditions to a nucleotide sequence selected from the group consisting of: SEQ.ID.NO.:1, SEQ.ID.NO.:2, SEQ.ID.NO.:4, and SEQ.ID.NO.:28.
20
5. An expression vector comprising the DNA of claim 1.
6. A recombinant host cell comprising the DNA of claim 1.
25
7. A CG1CE protein, substantially free from other proteins, having an amino acid sequence selected from the group consisting of SEQ.ID.NO.: 3, SEQ.ID.NO.:5, and SEQ.ID.NO.: 29.
30
8. The CG1CE protein of claim 8 containing a single amino acid substitution.
9. The CG1CE protein of claim 9 where the substitution occurs at position 6, 85, 93, 227, or 299.
35

10. The CG1CE protein of claim 9 where the substitution is a conservative substitution.

5 11. The CG1CE protein of claim 8 containing two amino acid substitutions.

10 12. The CG1CE protein of claim 8 containing an amino acid substitution where the substitution does not occur in a position where the amino acid present in CG1CE is also present in the corresponding position in one of the *C. elegans* proteins whose partial amino acid sequence is shown in Figure 7.

15 13. An antibody that binds specifically to a CG1CE protein where the CG1CE protein has the amino acid sequence selected from the group consisting of SEQ.ID.NO.:3 and SEQ.ID.NO.:5.

14. A method of diagnosing whether a patient carries a mutation in the CG1CE gene that comprises:

- 20 (a) providing a DNA sample from the patient;
(b) providing a set of PCR primers based upon SEQ.ID.NO.:2 or SEQ.ID.NO.:4;
(c) performing PCR on the DNA sample to produce a PCR fragment from the patient;
25 (d) determining the nucleotide sequence of the PCR fragment from the patient;
(e) comparing the nucleotide sequence of the PCR fragment from the patient with the nucleotide sequence of SEQ.ID.NO.:2 or SEQ.ID.NO.:4;

30 where a difference between the nucleotide sequence of the PCR fragment from the patient with the nucleotide sequence of SEQ.ID.NO.:2 or SEQ.ID.NO.:4 indicates that the patient carries a mutation in the CG1CE gene.

35 15. The method of claim 15 where the DNA sample is genomic DNA.

16. The method of claim 15 where the DNA sample is cDNA.

5 17. A DNA or RNA oligonucleotide probe comprising at least 18 contiguous nucleotides of at least one of a sequence selected from the group consisting of: SEQ.ID.NO.:1, SEQ.ID.NO.:2, SEQ.ID.NO.:4, and SEQ.ID.NO.:28.

10 18. A method for determining whether a substance is an activator or an inhibitor of a CG1CE protein or a mutant CG1CE protein comprising:

(a) recombinantly expressing CG1CE protein or mutant CG1CE protein in a host cell;

15 (b) measuring the biological activity of CG1CE protein or mutant CG1CE protein in the presence and in the absence of a substance suspected of being an activator or an inhibitor of CG1CE protein or mutant CG1CE protein;

20 where a change in the biological activity of the CG1CE protein or the mutant CG1CE protein in the presence as compared to the absence of the substance indicates that the substance is an activator or an inhibitor of CG1CE protein or mutant CG1CE protein.

FIGURE 1A

```

1   ccaaaaaaatt gttctcttgg ggggttggggc gacaagcggg aaggaggaggc
51  attttgggca aattggctta ttgccacgca agggctttta caccttaggt
101 tgggtgggttc acaggttgca ggcaaccac catggcacac gtatacctat
151 gtaaccaacc tgcaccatca tgtataccta tgtaaccaac ctggtacatt
201 ctgcacacgt atcccaggac ttagagtga aaaaaaagt ggtgtgtaga
251 aaaatcacct gcaatctcag catagttaac gcttagtaca tttcagagag
301 agaggggtgac aggaaggga ggatgagagt gggtttaaga cacaagggtca
351 tattataaaa tcagggcttc tggaagtta gtcccaaac cacacatctc
401 ataatcccc tgcagtgttg attaaaatgc aacatcccta aggccacaga
451 ctgagactct ggagaaagat ccagaaaact gcccgttta taaacatttg
501 ggcgattctt acggcctcta aagaccaaga accactgctg cctagagctc
551 tgctctcttc attgaacaat acaagaggag tgtgtaggta gacaccacc
601 acttccaaca gcttaggaga gcccttgagt atggattgat gtattaaaat
651 ttattgaatc acatgtgag attttacca gctgcccgtg gggatctggg
701 catttattcc catattgcac tggctggctg gaagccagca gcataaactc
751 cagggctgtt ctgtcaacct ccaccagact cacccccctc caccagcccc
801 ggcaggcttc tccttccatc tctctgaagc aacttactga tgggccctgc
851 cagccaatca cagccagaat aacgtatgat gtcaccagca gccaatcaga
901 gctcctcgtc agcatatgca gaattctgtc attttactag ggtgatgaaa
951 ttcccaagca acaccatcct tttcagataa gggcactgag gctgagagag
1001 gagctgaaac ctaccgggg tccaccacac caggtggcaa ggctggggacc
1051 agaaaccagg actgttgact gcagcccgtt attcattctt tccatagccc
1101 acagggctgt caaagacccc agggcctagt cagaggctcc tccttcttgg
1151 agagttcttg gcacagaagt tgaagctcag cacagcccc taacccccaa-
1201 ctctctctgc aaggcctcag gggtcagaac actgggtggag cagatccttt
1251 agcctctgga ttttagggcc atggtagagg ggggtgttgc ctaaattcca
1301 gccctggtct cagcccaaca ccctccaaga agaaattaga ggggccatgg
1351 ccaggctgtg ctagccgttg ctcttgagca gattacaaga agggactaag
1401 acaaggactc ctttgtggag gtccctggctt agggagtcaa ctgacggcgg
1451 ctgagcactc acgtgggcag tgccagcctc taagagtggg caggggcact
1501 ggccacagag tccCAGGGAG TCCCACCAGC CTAGTCGCCA GACCTTCTGT
1551 GGGATCATCG GACCCACCTg gaacccacc tgtgagtaca aggtgcccc
1601 ggtggactgg gctggggctt tgaggccttc aggggttgat ggccatcttg
1651 cgtattttgtg tgggatatgc acacacaggc agcacatgcg caggtgtgtg
1701 ggcacctgtg tgtctgtgca aatgccctga ggtgggaatg agcttgggtg
1751 gcatcaggag cgacagccag ccagtgtggc tgcagcaaaa cacacaggga
1801 aagaatggag ggggcatcaa tcaactgaca aattatttat agagctcccc
1851 ctaaaaaaa gaaggtctct tctttcgata gaagaaggga gagaggggg
1901 ttgtccttat aaatataagg gaggagccgc ccctcaaaa ataaggagg
1951 gaggaccaa gaccccgagg gttgtgtgtt ttccaggggg agctcgaacc
2001 ctttagaggg agcgtgggag aaccgctgta ttcaggcctc tcgagagaaa
2051 aggagcggcc gcccaaaaaa tatccctccc gggcgataag aaatgggtggc
2101 ctctctcaaa aagatgaaga ggaagccgga gttgtatgtg ttgatatttt
2151 taaaactcca ggtagnnnnn nnnnntgctt cagttaaattt ttattgagcg
2201 ccttctacga gaacacaaga ggagcttcca ttctgaggag gaaacaggca
2251 ggaaacaggc agatatcctg tataatttca agtagtgata agtgctctct
2301 agaaatatca agcaaggta ggagacacag agcaccgggtg gcagtggggc
2351 tctatttcca ggttggatgg ttgggaacat cctttctaaa gggaaacctgg
2401 agtgggaagg aaccatgcag gtatctcagg aagagcttcc tccaggcagg
2451 aagatcagca ggtggaaagg ccctggagcc accattcagt aaacatcatt
2501 tgagcatctc taccagctag gttccattat gggaaatggga atatgggtgg
2551 ggacagggct gcctgggtccc ttccatactt ctcacactag ggtgggttag
2601 agagcttggg agctaacgaa caagatgggc tgagaacact gcctagccca
2651 gaggacctga gcttagtgtg tagacattgc tgctgttact gcctttgtcg
2701 ttgtattatt tatttattta tttattgatc ttaagacaga gttttgctct

```

FIGURE 1B

2751 tcttaccag gcttgagtgc aatggcgtga tctcagctca ctgcaacctc
 2801 cacctcctgg gatcaagcga ttctcctgcc tcagcctcct gagtagctgg
 2851 gattacaggc acccgaccca cgcctggata atttttttgt attttttagta
 2901 gagacagggt ttcacatgt tggccaggct ggtctcgaac tcctgacctt
 2951 aggtgatcca cctgcctcga cttcccaaag tgctgggatt ataggcatga
 3001 gccactgcgc ccagtgatta tagaaagtta aaggcacatg gcaatgcaca
 3051 cgcctatcta cgtcttcctt gccaaagcaa agggcagcct ctgggctcac
 3101 tttcttgctg ttctacttcc aaaaggcagt cagaactggc agggccttgg
 3151 agaccacttc atccacctcc tagggctcct atgggagagt tgaggctccag
 3201 agcaggggaag ggtcctgaca ggctctgacc agggcctctg atccctacaa
 3251 accccaatc ggtgtccctc tctaccagGA CCCAAGCCCA CCTGCTGCAG
 3301 CCCCTGCTT GGCCA TGACC ATCACTTACA CAAGCCAAGT GGCTAATGCC
 3351 CGCTTAGGCT CCTTCTCCCG CCTGCTGCTG TGCTGGCGGG GCAGCATCTA
 3401 CAAGCTGCTA TATGGCGAGT TCCTAATCTT CCTGCTCTGC TACTACATCA
 3451 TCCGCTTTAT TTATAGgtaa agctggcagg gctgggcccgg ggggcctggg
 3501 aaggatgtgg ctggggctgg gagctgggag ctctggggg cctcccagcc
 3551 agctcagggc ccagtgcacc agtccactac aacactaagc tgggctcctg
 3601 accagctcct gggcactgga gctgaggctg cgcgctgggg gctgggcaga
 3651 gtaaagaagt cacactgaga ggctgctcaa gccaggccag cagggtttta
 3701 gccacccttc ctccaacccc agggaggacc ctggagccca ggctttgtct
 3751 ggccccactc tactggcctg ttttactgaa tcccacacag actcataggc
 3801 ccacatagta cattaataaa gagagagaga gagagagaga gagagagatg
 3851 gagtctcact gtgttgtcca ggctgggtctc gaactcctag gctcaagcaa
 3901 tccccctgcc ttagcctccc aaggggctgg gattacaggt gtgagctact
 3951 gcacttgacc aaccacatgg tacttttttt ttttttttt ttttttgaga
 4001 cagggtttca ctccatcacc caggctggag tgcagtgggg gcaatcttgg
 4051 ctactgtaa cctctgcctc ccagggtgcaa gcgattctcc tgccttagcc
 4101 tcttgagtag ctggaattat aggcacacac caccacgcct ggctaatttt
 4151 tttttttttc tgtattttta gtagagacag ggtttcatca tgttgccag
 4201 gctggctctg aaccctgac ctcaagtgat ccacccacct cggcctccca
 4251 aagtgtctgg attacaggtg tcagccacca tgcacagccc acatggtaca
 4301 ttttttaaaa ttatttttta attaaaatgt ttatctaagg ccagtagcag
 4351 tgactcgcgt ctgtaatccc agcactttga ggggccaagg tgcggggatc
 4401 acttgagcct gggagttcag cgtgggcaac atagttagac cccgtctcta
 4451 ccaaaaattt aaaaaattag ctgggagtgg tggcatttgc ctgtgttccc
 4501 agctacttgg gaagctgagg tgtggggatg gctgaagcct gtgaggtcga
 4551 ggctgcagtg agctatgatc acaccactgc acttcagcct gtagtgcagg
 4601 ctatctcaaa agcaaacaaa ataagtgtta tctaaacggt aaggtataat
 4651 cacagaatat atgatagcat tttaaattga aaaagcatta atgattacat
 4701 ggattgtaaa atatcaaata catgaaatcc ttgtgttctt aataatgcta
 4751 gcaacaaggc acatttggtt tttactaggg caccaaggta ctttaaaaaa
 4801 agttaggggc agccacagg gctcacacct gtaatcccag cactttggga
 4851 ggccaaggca ggaggatcac ttgagcccag gagtttagga cctgagcaac
 4901 ataggagat cctgatcttg tctctataaa aaattaaaaa attggctagg
 4951 ccttttggt tacacccgta atcccagcac tttgggaggc cgaggcgggt
 5001 ggatcatgag gtcaggagt caagaccagc ctggccaaca tagtgaaccc
 5051 aatctctact ataaatacaa aaattagccg agtgggggtg cagcacctg
 5101 tagttccagc tactcaggag gatgaggccg gagaatcgt tgagccggg
 5151 aggcagaggc tgcagtgagc cgagaccatg ccattgcact ccagcctagg
 5201 tgacagagt agactccgtc ttaaaataat attaaaatct taaaatgatc
 5251 tgggcatggt ggcttatgcc tgtagtccca cccagctctt caggaggctg
 5301 aagcgggagg attgcttcag cccaggaggt tgaggctgca gtgagtcag
 5351 actgtgccgc tgcccttgag cctgggtaac agagcaagac cctatctcaa
 5401 aacaaacaaa caaacacaaa aacaaacaaa aaccaataaa ccaaaaacat
 5451 ttatctaaac aataaaataa aggacagata taatcaccga atatatgata

FIGURE 1c

5501 gcatttttaa ttgaaaaagc actaatgact acaatggatt ataaaacatc
5551 aaatacataa aattcttaag ttctctctaa taccaaatatc aaagcacatt
5601 ggtcttttgg ttttacttgg gcaccaatgc atgctgaaaa agagtcgttc
5651 atttttttaga gtagtttttag gttcacagca aaattgagca gaaggtagag
5701 ttctcatgtg tctcttttget cctccccctg cccccagcct cccactatc
5751 aacaccccca cactacagtg gtagatttat tacaatccct gaaccacag
5801 tgacacatca ctatcaccca aagttcatag cgtacagcag ggttcactct
5851 tgggcagtac attccatggg tttggataaa tgtgtaatga tgtctccacc
5901 atcacagcat caggcagagt agtttctactg ctctaacaaa atcctctgcc
5951 tattcacccc tctcattaaa gccaaacact ctgtttcctt tttctcttt
6001 agagacagtg tctcgtctg tcaaccaggc tgaagtgcaa tggcaatcac
6051 agcccatgtc agcctccaac tcctgggctc aagtgtacct cctatctcag
6101 cctccagtgg ctacgactgc aggcatacgg caacggcacc caactaattt
6151 tttgtagaga tagggctctg ctatgttgcc caggctgggc ttgaactctt
6201 ggtcctgct tagcctccca gagctctggg attacaggcg tgaaccaccg
6251 tgcccgctcc aaacactctg tttcgacctg cttttaaaca actgaccctt
6301 ggatgcattc aaaggatcag ggtgtctgaa actggcctct gcagcaggac
6351 cttccttctt acacatctcc cagtggccag tgtgaggatt ctccccacaa
6401 gaaaccactg gagggggcct cctcctgtcc ggggttgggg ctgtacaagg
6451 agcatcatgg acctggctca ggccctcagga ggggcccctg gctgggaaa
6501 atgtgggata gcacagaggc agtcccactc ctaccagggg ccgggctaga
6551 cctggggaca gtctcagcca tctcctcgtc gcgtccacac aattccaccc
6601 ccacccccac cccagGCTG GCCCTCACGG AAGAACAACA GCTGATGTTT
6651 GAGAACTGA CTCTGTATTG CGACAGCTAC ATCCAGCTCA TCCCCATTTT
6701 CTTCGTGCTG Ggtgagttcc ccttcttggc tgttccgggt ccctgtggcc
6751 gccaggctc cagacaggcc aggggaggat cacgaggagc tgcggcaagg
6801 ggctggggag ggggcggggg aacgccagcg gcaggtcggc gcctctctgt
6851 agggaaaagg gcgactgca gccagagaaa ctgaagttag acgttaggta
6901 agacgtcctg ccgttagcaa tgaaaacccc attttctgag ggaagcgtg
6951 acatcatggt ccctggagcc cctgcgcggg aggggagggg gtctggcgga
7001 tttctgggac cagcaggggg acccccgggt gacagaacct ttggggctct
7051 cgcgctcca tgcgaggctc tgctgcctc tcgtcccgga gcgccttcca
7101 ggagggctgg gggctaggcc cgctcgcagc agaaagctgg aggagccgag
7151 gcacgcggg gcgctgggccc ctgggctctg gccgcagcct ggcccctcgc
7201 ccctcgcccc ccgcccctcc tgcccagGCT TCTACGTGAC GCTGGTCTGTG
7251 ACCCGTGGT GGAACCACTA CGAGAACCCTG CCGTGGCCCG ACCGCTCAT
7301 GAGCCTGGTG TCGGGCTTCG TCGAAGGCAA GGACGAGCAA GGCCGGCTGC
7351 TGCGGCGCAC GCTCATCCGC TACGCCAACC TGGGCAACGT GCTCATCCTG
7401 CGCAGCGTCA GCACCGCAGT CTACAAGCGC TTCCCAGCG CCCAGCACCT
7451 GGTGCAAGCA Ggtgggcgga ccgggagcaa cggggaggca ccgggcagag
7501 ccagggggccg agatgggcgc ggcaggaacg gaagatgggt ggagccaaag
7551 tccccggac tcgggggact gggtaggagc aggagtgggg tgtggtcaag
7601 atttgggggt ccaattgggc gggacagagt cgggtgtctg aaggtggggc
7651 gagggcagga gccaccctc cgagagtagg agtctgaggc agggctaagg
7701 acccttgagg gataatggaa agaagggtga cggcttggga actggtgagg
7751 tactagggtc tacttccctc tgcccttgcc cctcttgatc tccggtttcc
7801 actctggagg tatgggacat tgggtctctga caccctctca gcctggcctg
7851 acctggctct ggttaataag acagaccag gctaggcggt gtggctctcg
7901 cctgtaatcc cagtgttta ggaggcaaag gtgggaagat cgcttgagcc
7951 cagctgtttg agacgcccct gagcaacata gcgagacccc catctctaca
8001 aaaacattaa aaattagcag ggcattgggt cgtgtgcctg tagtctgagg
8051 ctgagtatcg ggaggctgag gcaggaggat cacttgagcc cagagttcc
8101 aggctgcagt gcgctaagat cgcaccgctg cactccaacc tcggtgacag
8151 agccagaccc tttctctgga aataaataaa taccctgccc acatgctcag
8201 cccagaacag cacctagtag gtgctcagaa atttttttgt tgttgaaaga

FIGURE 1D

8251 aagaggatgg caaaggagtg ctgagggtcc tatagggtcag caggtgccgg
8301 ccatcccttc tgcagggtct cccacccacc gccttcttca ctccactctg
8351 cagGCTTTAT GACTCCGGCA GAACACAAGC AGTTGGAGAA ACTGAGCCTA
8401 CCACACAACA TGTCTGGGT GCCCTGGTG TGGTTGCCA ACCTGTCAAT
8451 GAAGGCGTGG CTGGAGGTC GAATCCGGA CCTATCCTG CTCCAGAGCC
8501 TGCTGAACgt gagccactg tacagacagg gctgccgcag agtgggaagg
8551 gttgtggtcc acaggaaaca aggtttccta caaagagaag ccttgggccc
8601 ctgagggtct tccgagagcc ggagggtggg ttgcagaatc ttttccaaca
8651 gcaatccaca gcccagagtg gtcccttctc agaggccctt cctctctctc
8701 caagtctgtg aggtcctggt tcccttttga tagatgagga agctgagaca
8751 caaagagggt tagtgagctt cccatggcca cacagccagg aatggacat
8801 aggtaccagg ccttgggtacc tggagaagag gtggggggcga gccaggggtg
8851 ggggcagggt gtgttcagaa ccccatcccc ctctctctgccc cccagGAGA
8901 TGAACACCTT CCGTACTCAG TGTGGACACC TGTATGCCTA CGACTGGATT
8951 AGTATCCAC TGGTGTATAC ACAGgtgagg actaggctgg tgaggctgcc
9001 cttttgggaa actgaggcta gaaggaccaa ggaagcagct ggggtgggaa
9051 gggctcacct agaggctaag tggctcccc gggagttggg tccacacttt
9101 gaagtgggt ctggactttg aagtggcaag ttctaagagt ccaggctcct
9151 gcctggccca gtccagtaga ggcaatgtga ttatcccat attaaagaga
9201 ggttggccgg gcacagtggc tcatgcctgt aatcccagca ctttgggaag
9251 ctgaggcagg tggatcacct gaggtcagga gttcgagacc agcctggcca
9301 acatggtgaa accccatctc tactgaaaat acagaattag ctgtgtgtg
9351 gtgcacgcct gtaatcccag ctacttggga ggctgaggca ggagaatcgc
9401 ttgaaccggg gaggtggagg ttgcagttag ctgagatcat gccactgcac
9451 tccagcctgg gcgacacagc aagactctgt ctcaaacaaa caaacaacaa
9501 aacaacaaa caaacaacaa aaggggttaa cagagccctt aagtcacata
9551 agtgtgcaag tcagaacaag gccttggctc cctgtctcag actcccagcc
9601 cctggagcat cctgatttca gggttccac ctagccctt gctaccacat
9651 cctcctctc ctccctctcc tcccagGTGG TGACTGTGGC GGTGTACAGC
9701 TTCTTCTGA CTGTCTAGT TGGGCGGCAG TTCTGAACC CAGCCAAGGC
9751 CTACCCTGGC CATGAGCTGG ACCTCGTTGT GCCCGTCTTC ACCTTCTGTC
9801 AGTTCTTCTT CTATGTTGGC TGGCTGAAGG TGGGCTCTC CAGGGCCCTG
9851 CTGGGCTGGA GGCATGGCCA GAGGGGTCTC GCGCAGCAGC TGCTTGAGAC
9901 GAGGATGCAG TGTGAGGAAA GGAAGGTCTC ACGGGTAGAA AGCAGCCAGG
9951 CGTGGTGGCG CACACCTGTA ATCCAGCTA CTGGGAGGC TGAGGCAGGA
10001 GAATCGCTTG AACCCGGGAG GCGGAGGTTG Tggtgagttg agatcgtgcc
10051 actgcactcc agcctgggca aaagaatgaa actctatctc aaaaacaaca
10101 acaacaacaa aacaaagccc taagggtcag aagccctgc cctttagaag
10151 cagagcgaac actctctat taagatgctg ttgggtgtct ttttactca
10201 gtagctgtcc agtattctcc acacagcata atcgacagat tctaatacaa
10251 atttcttcaa ctcttaatte ctctttgtg ccaccatttt ttctctacc
10301 tectaattta tgaatgggtt agtatgctct gcttctgcat tgagacaaaa
10351 tacagagaga gagaaagatc tatcttaate ccgccccatt ttagttggaa
10401 aaaaacttta ttaaatcagg caagtaaaat ccgccaagga ttgnnnnnnn
10451 nnnagatggt ctgaatcaga gagttttctc tcgagctctt tatcttctc
10501 tcttctgtt gccacccac tctctctccc ttctacctt cctttatttt
10551 ttggtaattg ggggtgaagt ctctgtctct gcccttctg tcaactgtgac
10601 acacacacac acacacacac acacacacac acacacacac attcctattc
10651 ctctaaattc cccctgcacc cccagttatc ttgggtttct gcagatcaaa
10701 acaaatcaca cttttatgct tgaaattctc cagggtgccc cagtgccctg
10751 caagatgtcc cctggacccc taaggcagac gcgtgtcacc tcttcggggc
10801 tttgttaggg catttttagag gttgctatcc aggaatctgc ccacctagac
10851 tgccctttag ttcagccag cttcagtata tatctctggt gcatgaatga

FIGURE 1E

10901 ataaaattat gcaactccag gtaagatata tgagggtgaga taaaggcagt
10951 gactcagccg agtgatacac tcagggacag ctgtgggtgt tcaggggaagg
11001 actggctcag aagagttaga ggggctgtgt ccagaagtgt gtgggtgcct
11051 acaagtgtgg ggggctggag ccctaaactc tgcctttgaa gacagtggtc
11101 aggcaggaag ggcttcattg ggtgtggaaa tagcagcagc tgagggttaa
11151 agggggaagc tggctttgag gagttctgcc tgagggttta cagagcctca
11201 cctgtcccca agGTGGCAGA GCAGCTCATC AACCCCTTTG GAGAGGATGA
11251 TGATGATTTT GAGACCAACT GGATTGTCGA CAGGAATTTG CAGgtatggg
11301 gagagggaga gaaaccatac catggacctt ccccaaagt gacccaaaga
11351 gagtcctcc ctcctgcagc cagtattca ctcacaggat tctcacctca
11401 atctttgagg ctgcaggcag gcacccatct cccatttca caggcagggg
11451 aactgaggtc cagagagagg gagagattcc tccaagtcac caggcacata
11501 caaggctctg cctgggatga tctttctgtg ggacttcttc tgtccctggt
11551 gaccagGTGT CCCTGTTGGC TGTGGATGAG ATGCACCAGG ACCTGCCTCG
11601 GATGGAGCCG GACATGTACT GGAATAAGCC CGAGCCACAG CCCCCCTACA
11651 CAGCTGCTTC CGCCAGTTC CGTCGAGCCT CCTTTATGGG CTCCACCTTC
11701 AACATCAGgt gtggccagag ccagggggct ggggtgggaag cccctcctag
11751 tgcaggggtc tgcctaggaa cttagaatag cactagttaa tgcatacagg
11801 ttgcttcagt aagtgtcagg cactgtacta tgctctttat aaacattaac
11851 tatttttttc ctcccaataa ttctggtttg ttatcccaag ttttcagata
11901 attaaagtac aggttcagag agagtaagtt gtccaaggcc acatagctac
11951 caaatggtgc atttgctact cgaaggacag cctatgatca gtgatgcagt
12001 ggaacgttag gacctggctc ttgtcatcca gaactatgtt ttcttttctt
12051 tttgagacag tatctcgtct tgtcgcccag gttggagcgc agtggcggtga
12101 tcttggctca ctgcaacctc cgctcctggt gttcaagtga ttctcctgct
12151 tcagcctccc cagtagctgg gattacaggt gccacacacc acaactggct
12201 aatttttcta cttttagtag agatgaggtt tcaccatgtt ggccaggtg
12251 gtctccaact cctgaccagt aatctgcccg ctttggcctc ccaaatgtg
12301 ggaattatag gtgtcaaaac tatgttttct gataagctac gatgcttga
12351 tgggaagtgg aagtggggtt ccctgggatg ggggaggggc agcaaagtcc
12401 cagcaggcag ccaggccatc acaggtacct cctgaattga ctttgtccta
12451 ccgagtaaag ggctcaggcc acccacagca gccagactta tccccacatg
12501 gtcccacttc cctgattcca tctgaatccc tcttgagctg cagtgggctg
12551 aagggtatc ccagctggct ctttctcccc aggacaacag agttgaaagt
12601 gccttggaga gtgttgggca catgtcaggg ttcatactca aggggttctt
12651 ccacggatc cagtgtgtt ctgcctgtt ctttctttt ttttttttta
12701 aacggagttt cactctgtt gccagagct ggagtgcagt ggcataatct
12751 cggctcactg caacctccgc ctcccagatt caagcaattc tccctgctca
12801 gcctcctgag tagctgggat tataggtgcc agccaccaag cccggctaat
12851 ttttgtattt ttagtagaga cagtttcacc atgttggcca ggctgggtctc
12901 gaactcctga cctcagggtga tccacctcc tcagcctccc aaagtgtgtg
12951 gattacatgt gtgagccact gtgcctggct gcttgttctt ttaagaacca
13001 aatatcctac tagactgcaa tgcagttaa ctacagtcta tagatactgt
13051 gaggaatggt tgggaaggct atcaaatgaa ggctggaggc ttgcttaggt
13101 cagaaacatt tctggaggat gactttgagc cctacatggt ctgtacccca
13151 gcagctgaag gttgttgagg gatggggagg gctgaaaaca gaacgataaa
13201 gcatagacct tgtctccaag gaatgcacaa tttatggagg gagctcaaac
13251 ccaagtctca aactctggat acaagggtaca aagtactgga tgtccagaaa
13301 agggacagaa catggaacac agtcatctt gtctgcctgg gaggcggctt
13351 ccagctgggt ctggagctga gccatggaac atgggaagaa tctgaacttg
13401 ggcaagggca ggccatactc tctggttagat aagctttcct tgcagggtaa
13451 aggtctgggg ctcccgggat gcctgttgc aggaagtcaa atttctctt
13501 gtggatgtca ctcccagttg gaaccacaaa ttcctggcat tgcacagagt
13551 cactcatggg cctcatctga accactcatg ccagggcacc agtgtttctg

FIGURE 1F

13601 actgcctgga gtgaggggtt ttacagggga agtgaatgat gaggaggcct
13651 ttacacgcca ggcgggggtg ttgcgggggt tggatgttaa ctctgggtcaa
13701 gagggaatca acaaacagtg aggtgagctg ggcctggagg gatcacgggg
13751 aggtacagta cagatcagga gagaggtgag agctggggca tgggtgaggaa
13801 gacgggtgtg ccttggcttg ggccaactga gagagaggag cgggggtaag
13851 ggagaagtaa ggccaggtgt tggtcctttg tccactggct cagccctgca
13901 tctcctgttt ctttccagCC TGAACAAAGA GGAGATGGAG TTCCAGCCCA
13951 ATCAGGAGGA CGAGGAGGAT GCTCACGCTG GCATCATGG CCGCTTCCTA
14001 GGCCTGCAGT CCCATGATCA CCATCCTCCC AGGGCAAAC CTGCCCCAAA
14051 ACTACTGTGG CCCAAGAGGG AATCCCTTCT CCACGAGGGC CTGCCCCAAA
14101 ACCACAAGGC AGCCAAACAG AACGTTAGGG GCCAGGAAGA CAACAAGGCC
14151 TGGAAGCTTA AGGCTGTGGA CGCCTTCAAG TCTGCCCCAC TGTATCAGAG
14201 GCCAGGCTAC TACAGTGCCC CACAGACGCC CCTCAGCCCC ACTCCCATGT
14251 TCTTCCCCCT AGAACCATCA GCGCCGTCAA AGCTTCACAG TGTACAGGC
14301 ATAGACACCA AAGACAAAAG CTAAAGACT GTGAGTTCTG GGGCCAAGAA
14351 AAGTTTGA TGTCTCTCAG AGAGCGATGG GGCCTTGATG GAGCACCCAG
14401 AAGTATCTCA AGTGAGGAGG AAAACTGTGG AGTTTAACCT GACGGATATG
14451 CCAGAGATCC CCGAAAATCA CCTCAAAGAA CCTTTGGAAC AATCACCBAAC
14501 CAACATACAC ACTACACTCA AAGATCACAT GGATCCTTAT TGGGCCTTGG
14551 AAAAACAGgtc tgtcctccac ctgaaccagg ggcaactgcat tgccctgtgc
14601 cccaccccag cttecccttgc tctgagccta ccttctctcc acaatttctt
14651 aggggttccat cactgccaga gcacactgga cctacgcca gcactggctt
14701 ggggatatata cttggccacc ttcacagggga tccatagggaa gtgttcggga
14751 cctttttctca cttcacccctg gtatcacccg gaagacttct tgggaccagg
14801 tgaaggaaga tgaggttgtg ctgaccagaa tgctgctgga gaactgcccc
14851 agggctgaca ggccaggtt agctgagcag atgttatcac tggccccaac
14901 ttactttgag caagggtggc tgacccaaaa ccatgaggtg gcagtcagct
14951 ggatgacaga tgaacacttc ccccataact atttagggta gtacccaagc
15001 actacaggaa aggggtggcag gaactgcctc actcctagga actggtagat
15051 ggtgaggttg aggggtgtcca gcgcccttag gtcattttct cactgcctgg
15101 gaacctcacc aaaatacttc ttgttctctt ggggtcagcc caaagctgtc
15151 acaaaatcag atatttccct ttattccaga tttcctggac actgtcacc
15201 aattataaac accccacttc agccccaatc acgtgggagg aagtgttaact
15251 tcccttttct ggattctcaa gcagttactt tcacgggtca gaacacgcag
15301 ctattatgat tgaacctta aaagggcaac aatttcantc ttgttctag
15351 gctaagacag gaacttggca aacatctgtg gcctgttcag caaaggatgt
15401 tcataatttaa gaatcttgtc ttgggtctgg tgtggaggca agtgaatcac
15451 aggaggtcag gaggtttgaga ccaacctggc caacatgatg aaaccccatc
15501 tctacaaaaa aaaatacaaa tcagctggcc gtcgtggtgt gcctgtagtc
15551 ccaacgcagg aggttgagg gagaattgct tgaacccagg aggtgggtgt
15601 tgcatgaga ttgagcaact gcaatccagc ctgggagcag gactgagact
15651 gtctcaaaaa aaaaaaaaaa aggatcgtct caacctttgc cctcctactg
15701 caacattttg gtatttgaaa tgaaggtacc ttccatactt atgctgttaa
15751 tactttcatt ctcactagGG ATGAAGCACA TTCCCTAACCT GCTTCCTAAT
15801 GGGGATGCTT CGCCAGCCAG GTCCTCACCT GTGTGTACAC CAGCAGGACA
15851 CTGATCCAGT CACAGCCATA CAGCTGTCCA CACTGAAGAA CGTGTCTTAC
15901 AACAGCCTGA ATCAAAATGGT TAGCTTAATA GATAAAAAATC CCAGACTACT
15951 TCAGCCTTAA ATGCCTTTTA TTCATAAAAA CTGTGAAAGC TAGACTGAAC
16001 CATTGGAAGC ATTAACTCA GACTCTGGAT TCAGAGTCGG GAACCTTAG
16051 TTCTATCTGA ATCCAAGACA GCCACACCTT AGTATACTGC CCAAACCTAAT
16101 GAGTTTAATA AATACAAATA CTCGT (SEQ.ID.NO.:1)

FIGURE 2

CAGGGAGTCCCACCAGCCTAGTCGCCAGACCTTCTGTGGGATCATCGGAC 50
CCACCTGGAACCCACCTGACCCAAGCCCACCTGCTGCAGCCCACTGCCT 100
GGCCATGACCATCACTTACACAAGCCAAGTGGCTAATGCCCCGCTTAGGCT 150
CCTTCTCCCGCCTGCTGCTGTGCTGGCGGGCAGCATCTACAAGCTGCTA 200
TATGGCGAGTTCTTAATCTTCCTGCTCTGCTACTACATCATCCGCTTTAT 250
TTATAGGCTGGCCCTCACGGAAGAACAACAGCTGATGTTTGAGAACTGA 300
CTCTGTATTGCGACAGCTACATCCAGCTCATCCCCATTTCCTTCGTGCTG 350
GGCTTCTACGTGACGCTGGTCGTGACCCGCTGGTGGAACCAAGTACGAGAA 400
CCTGCCGTGGCCCGACCGCCTCATGAGCCTGGTGTGCGGGCTTCGTGGAAG 450
GCAAGGACGAGCAAGGCCGGCTGCTGCGGCGCACGCTCATCCGCTACGCC 500
AACCTGGGCAACGTGCTCATCTGCGCAGCGTCAGCACCGCAGTCTACAA 550
GCGCTTCCCCAGCGCCCAGCACCTGGTGCAAGCAGGCTTTATGACTCCGG 600
CAGAACACAAGCAGTTGGAGAACTGAGCCTACCACACAACATGTTCTGG 650
GTGCCCTGGGTGTGGTTTGCCAACTGTCAATGAAGGCGTGCTTGGAGG 700
TCGAATCCGGGACCCTATCCTGCTCCAGAGCCTGCTGAACGAGATGAACA 750
CCTTGCGTACTCAGTGTGGACACCTGTATGCCTACGACTGGATTAGTATC 800
CCACTGGTGTATACACAGGTGGTGAAGTGTGGCGGTGTACAGCTTCTTCCT 850
GACTTGTCTAGTTGGGCGGCAGTTTCTGAACCCAGCCAAGGCCTACCCTG 900
GCCATGAGCTGGACCTCGTTGTGCCCCGTCTTCACGTTCTTCAGTTCTTC 950
TTCTATGTTGGCTGGCTGAAGGTGGCAGAGCAGCTCATCAACCCCTTTGG 1000
AGAGGATGATGATGATTTTGAGACCAACTGGATTGTGACAGGAATTTGC 1050
AGGTGTCCCTGTTGGCTGTGGATGAGATGCACCAGGACCTGCCTCGGATG 1100
GAGCCGGACATGTACTGGAATAAGCCCGAGCCACAGCCCCCTACACAGC 1150
TGCTTCCGCCCCAGTTCCGTCGAGCCTCCTTTATGGGCTCCACCTTCAACA 1200
TCAGCCTGAACAAAGAGGAGATGGAGTTCCAGCCCAATCAGGAGGACGAG 1250
GAGGATGCTCACGCTGGCATCATTTGGCCGCTTCCTAGGCCTGCAGTCCCA 1300
TGATCACCATCCTCCCAGGGCAAACCTCAAGGACCAAACCTACTGTGGCCCA 1350
AGAGGGAATCCCTTCTCCACGAGGGCCTGCCCCAAAACCAAGGCAGCC 1400
AAACAGAACGTTAGGGGCCAGGAAGACAACAAGGCCTGGAAGCTTAAGGC 1450
TGTGGACGCCTTCAAGTCTGGCCCACTGTATCAGAGGCCAGGCTACTACA 1500
GTGCCCCACAGACGCCCCCTCAGCCCCACTCCCATGTTCTTCCCCCTAGAA 1550
CCATCAGCGCCGTCAAAGCTTCACAGTGTACAGGCATAGACACCAAAGA 1600
CAAAAGCTTAAAGACTGTGAGTTCTGGGGCCAAGAAAAGTTTGAATTGC 1650
TCTCAGAGAGCGATGGGGCCTTGATGGAGCACCAGAAAGTATCTCAAGTG 1700
AGGAGGAAAACTGTGGAGTTTAACCTGACGGATATGCCAGAGATCCCCGA 1750
AAATCACCTCAAAGAACCTTTGGAACAATCACCACCAACATACACACTA 1800
CACTCAAAGATCACATGGATCCTTATTGGGCCTTGGAACCAAGGGATGAA 1850
GCACATTCTTAACCTGCTTCTAATGGGGATGCTTCGCCAGCCAGGTCCT 1900
CACCTGTGTGTACACCAGCAGGACACTGATCCAGTCACAGCCATACAGCT 1950
GTCCCACTGAAGAACGTGTCTACAACAGCCTGAATCAAATGGTTAGCT 2000
TAATAGATAAAAAATCCCAGACTACTTCAGCCTTTAATGCCTTTTATTCAT 2050
AAAACTGTGAAAGCTAGACTGAACCAATTGGAACATTTAACTCAGACTC 2100
TGGATTGAGAGTCGGGAACCCCTTAGTTCTATCTGAATCCAAGACAGCCAC 2150
ACCTTAGTATACTGCCCAAACCTAATGAGTTTAATAAATACAAATACTCGT 2200
TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ.ID.NO.:2)

FIGURE 3

MTITYTSQVANARLGFSRLLLCWRGSIYKLLYGEFLIFLLCYYIIRFTY	50
RLALTEEQQLMFEKLTLYCDSYIQLIPISFVLGFYVTLVVTRWWNQYENL	100
PWPDRLMSLVSGFVEGKDEQSRLLRRTLIRYANLGNVLILRSVSTAVYKR	150
FPSAQHLVQAGFMTPAEHKQLEKLSLPHNMFVWPVWFANLSMKAWLGGR	200
IRDPILLQSLLNEMNTLRTQCGHLYAYDWISIPLVYTQVVTVAVYSFFLT	250
CLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGWLKVAEQLINPFGE	300
DDDDFETNWIHDRNLQVSLLAVDEMHQDLPRMEPDMYWNKPEPQPPYTAA	350
SAQFRRASFMGSTFNISLNKEEMEFQPNQEDEEDAHAGIIGRFLGLQSHD	400
HHPPRANSRTKLLWPKRESLLHEGLPKNHKAAKQNVRGQEDNKAWKLKAV	450
DAFKSGPLYQRPGYYSAPOTPLSPTPMFFPLEPSAPSKLHSVTGIDTKDK	500
SLKTVSSGAKKSFELLSESDGALMEHPEVSQVRRKTVEFNLTDMPEIPEN	550
HLKEPLEQSPTNIHTTLKDHMDPYWALENRDEAHS (SEQ.ID.NO.:3)	

FIGURE 4

CAGGGAGTCCCAACAGCCTAGTCGCCAGACCTTCTGTGGGATCATCGGAC 50
CCACCTGGAACCCACCTGACCCAAGCCACCTGCTGCAGCCCACTGCCT 100
GGCCATGACCATCACTTACACAAGCCAAGTGGCTAATGCCCGCTTAGGCT 150
CCTTCTCCCGCCTGCTGCTGTGCTGGCGGGGAGCATCTACAAGCTGCTA 200
TATGGCGAGTTCTTAATCTTCCTGCTCTGCTACTACATCATCCGCTTTAT 250
TTATAGGCTGGCCCTCAGGAAGAACAACAGCTGATGTTTGAGAACTGA 300
CTCTGTATTGCGACAGCTACATCCAGCTCATCCCCATTCTTCGTGCTG 350
GGCTTCTACGTGACGCTGGTCGTGACCCGCTGGTGGAACCAAGTACGAGAA 400
CCTGCCGTGGCCCGACCGCCTCATGAGCCTGGTGTGGGCTTCGTGGAAG 450
GCAAGGACGAGCAAGGCCGGCTGCTGCGGCGCACGCTCATCCGCTACGCC 500
AACCTGGGCAACGTGCTCATCTGCGCAGCGTCAGCACCGCAGTCTACAA 550
GCGCTTCCCCAGCGCCAGCACCTGGTGCAAGCAGGCTTTATGACTCCGG 600
CAGAACAACAAGCAGTTGGAGAACTGAGCCTACCACACAACATGTTCTGG 650
GTGCCCTGGGTGTGTTTGCCAACCTGTCAATGAAGGCGTGGCTTGGAGG 700
TCGAATCCGGGACCTATCCTGCTCCAGAGCCTGCTGAACGAGATGAACA 750
CCTTGCGTACTCAGTGTGGACACCTGTATGCCTACGACTGGATTAGTATC 800
CCACTGGTGTATACACAGGTGGTGTACTGTGGCGGTGTACAGCTTCTTCT 850
GACTTGTCTAGTTGGGCGGCAGTTTCTGAACCCAGCCAAGGCCTACCCCTG 900
GCCATGAGCTGGACCTCGTTGTGCCCGTCTTCACGTTCTTCGAGTTCTTC 950
TTCTATGTTGGCTGGCTGAAGGTGGGCCTCTCCAGGGCCCTGCTGGGCTG 1000
GAGGCATGGCCAGAGGGGTATGGCCAGCAGCTGCTTGAGACGAGGATGC 1050
AGTGTACAGGAAAGGAAGGTCTACGGGTAGAAAGCAGCCAGGCGTGGTGG 1100
CGCACACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCT 1150
TGAACCCGGGAGGCGGAGGTTGTGGTGGCAGAGCAGTCTATCAACCCCTT 1200
TGGAGAGGATGATGATGATTTTGAGACCAACTGGATTGTGACAGGAATT 1250
TGCAGGTGTCCCTGTTGGCTGTGGATGAGATGCACCAGGACCTGCCTCGG 1300
ATGGAGCCGGACATGTAAGTGAATAAGCCCGAGCCACAGCCCCCTACAC 1350
AGCTGCTTCCGCCCAGTTCCGTGAGCCTCCTTTATGGGCTCCACCTTCA 1400
ACATCAGCCTGAACAAAGAGGAGATGGAGTTCCAGCCCAATCAGGAGGAC 1450
GAGGAGGATGCTACGCTGGCATCATTTGGCCGCTTCTAGGCCTGCAGTC 1500
CCATGATCACCATCCTCCCAGGGCAAACCTCAAGGACCAAACTACTGTGGC 1550
CCAAGAGGGAATCCCTTCTCCACGAGGGCCTGCCCAAAAACCAAGGCA 1600
GCCAAACAGAACGTTAGGGGCCAGGAAGACAACAAGGCCTGGAAGCTTAA 1650
GGCTGTGGACGCCTTCAAGTCTGGCCCACTGTATCAGAGGCCAGGCTACT 1700
ACAGTGCCCCACAGACGCCCCCTCAGCCCCACTCCCATGTTCTTCCCCCTA 1750
GAACCATCAGCGCCGTCAAAGCTTCACAGTGTACAGGCATAGACACCAA 1800
AGACAAAAGCTTAAAGACTGTGAGTTCTGGGGCCAAGAAAAGTTTGAAT 1850
TGCTCTCAGAGAGCGATGGGGCCTTGATGGAGCACCCAGAAGTATCTCAA 1900
GTGAGGAGGAAAACCTGTGGAGTTTAACCTGACGGATATGCCAGAGATCCC 1950
CGAAAATCACCTCAAAGAACCTTTGGAACAATCACCAACCAACATACACA 2000
CTACACTCAAAGATCATATGGATCCTTATTGGGCCTTGAAAAACAGGGAT 2050
GAAGCACATTCTAACCTGCTTCTTAATGGGGATGCTTCGCCAGCCAGGT 2100
CCTCACCTGTGTGTACACCAGCAGGACACTGATCCAGTCACAGCCATACA 2150
GCTGTCCACACTGAAGAACGTGTCTTACAACAGCCTGAATCAAATGGTTA 2200
GCTTAATAGATAAAAAATCCCAGACTACTTCAGCCTTTAATGCCTTTTATT 2250
CATAAAAACCTGTGAAAGCTAGACTGAACCATTTGGAACATTTAACTCAGA 2300
CTCTGGATTTCAGAGTCGGGAACCCCTTAGTTCTATCTGAATCCAAGACAGC 2350
CACACCTTAGTATACTGCCCAAACTAATGAGTTTAATAAATACAAATACT 2400
CGTAAAAA (SEQ.ID.NO.:4)

FIGURE 5

MTITYTSQVANARLGSFSRLLCWRGSIYKLLYGEFLIFLLCYYIIRFIY	50
RLALTEEQQLMFEKLTLYCDSYIQLIPISFVLGFYVILVVTRWWNQYENL	100
PWPDRLMSLVSGFVEGKDEQGRLLRRTLIRYANLGNVLILRSVSTAVYKR	150
FPSAQHLVQAGFMTPAEHKQLEKLSLPHNMFVWPVWVFANLSMKAWLGGR	200
IRDPIQLQSLLNEMNTLRTQCGHLYAYDWISIPLVYTQVVTAVYSFFLT	250
CLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGWLKVGLSRALLGWR	300
HGQRGHGQQLLETRMQCQERKVS RVESQAWWRTPVIPATREAEAGESLE	350
PGRRLWWQSSSSTPLERMMMLRPTGLSTGICRCPCWLWMRCTRTCLGW	400
SRTCTGISPSHSPPTQLLPSSVEPPLWAPPSTSA (SEQ.ID.NO.:5)	

FIGURE 6

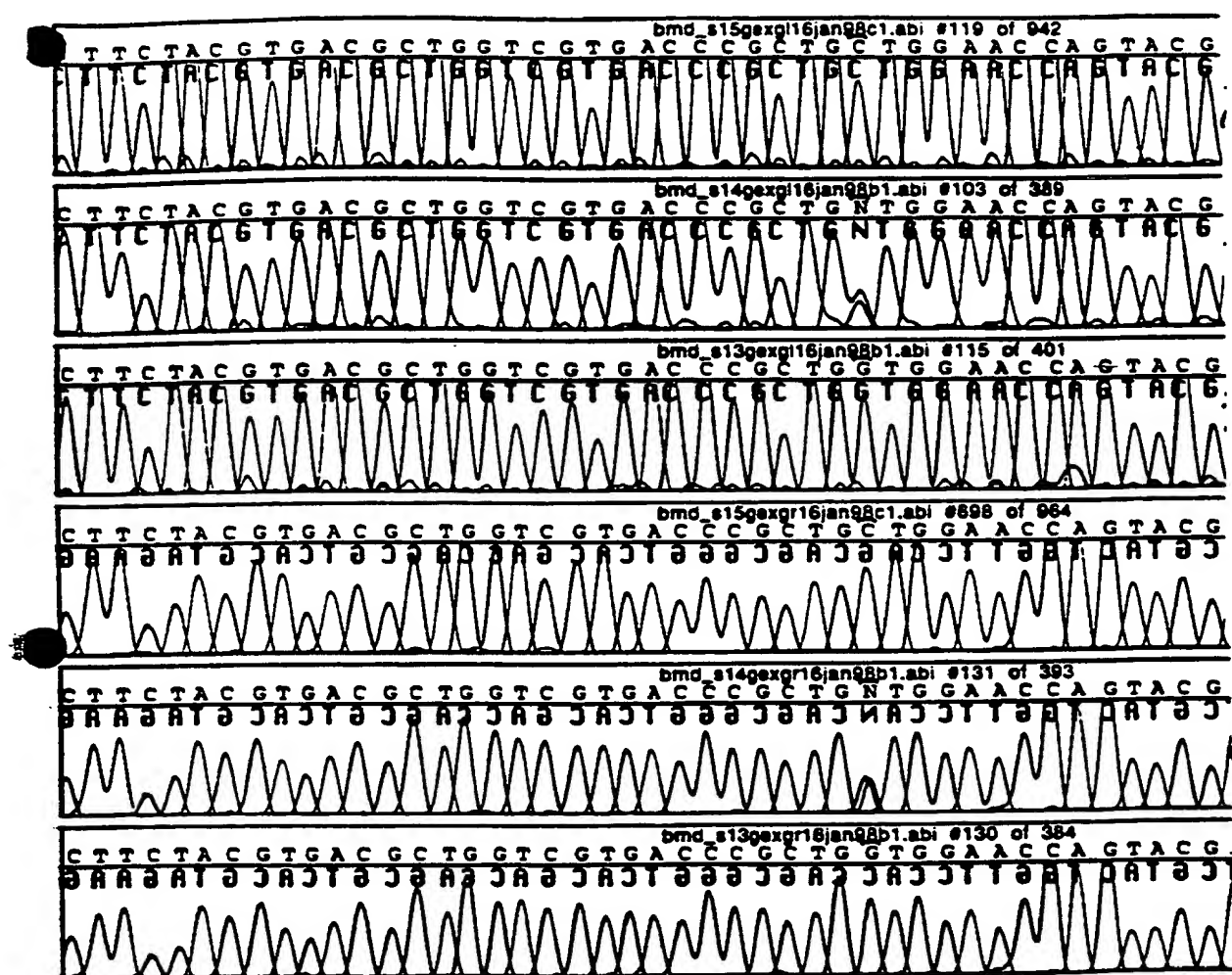


FIGURE 7

<u>GenBank/SwissProt accession numbers</u>	<u>Protein sequence</u>	<u>SEQ.ID.NO.</u>
CG1CE_protein	IPISFVLGFY VTLVVTRWYN QYENLPWPDR 2 (part)	
af016687 (PID:g2315833)	IPLTFMLGFY VTIIVGRWND IFLNIGWVDN 28	
z73105 (PID:e242363)	IPLTFMLGFY VTIIVRRWND IFANLGWVEN 29	
z73422 (PID:e244423)	IPLFVLGFY VTIVVDRWTK LWRTVGFIDD 30	
z73422 (PID:e244542)	IPLFVLGFY VTTVVNRWTK LYQTIGFIDN 31	
p34577	VPLDWMGLGFY IAGVLRREWY LYDIIGFIDN 32	
p34672	IPLNFMGLGFY VTAVVNRWTY LYQIIGFIDN 33	
p34319	LPLNFMGLGFY CNIIIRRWLK LYTSLGNIDN 34	
z68335 (PID:e217363)	IPINFMGLGFY VTTVINRWMT QFANLGMIDN 35	
z68753 (PID:e218704)	IPLTFLLGFY VSFVVARWGS ILNGIGWIDD 36	
af025458 (PID:e2429439)	IPVTFMLGFY VSIVYNRWTK VFDNVGWIDT 37	
u28412 (PID:g849242)	LPLTFMLGFY VTTVFERWRS ALNVMPFIES 38	
u70848 (PID:g1572760)	IPLTFLLGFY VSNVVSRAWR QFETLRWPED 39	
z81074 (PID:e351507)	IPLTFLLGFY VSNVVARWWR QFETLYWPED 40	
q09379	IPLTFLLGFY VAMIVRRWWD CCQLISWPDH 41	
z72509 (PID:e239377)	IPLSFLGFY VSLIVARWWE QFNCISWPDK 42	
z83221 (PID:e349023)	VPMQFMGLGY IGMVGERWGE SFENVSYIEK 43	

FIGURE 8A

1	GTGCCAAGCCATGACTATCACCCTACACAAACAAAGTAGCCAATGCCCCGCTCGGTTTCGTT	60
1	M T I T Y T N K V A N A R L G S F	17
61	CTCGTCCCTCCTCCTGTGCTGGCGAGGCAGCATCTACAAGCTGCTGTATGGAGAATTCCT	120
18	S S L L L C W R G S I Y K L L Y G E F L	37
121	TGTCTTCATATTCCTCTACTATTCCATCCGTGGACTCTACAGAATGGTTCTCTCGAGTGA	180
38	V F I F L Y Y S I R G L Y R M V L S S D	57
181	TCAGCAGCTGTTGTTTGAGAAGCTGGCTCTGTACTGCGACAGCTACATTCAGCTCATCCC	240
58	Q Q L L F E K L A L Y C D S Y I Q L I P	77
241	TATATCCTTCGTTCTGGGTTTCTATGTTACATTGGTGCTGAGCCGCTGGTGGAGCCAGTA	300
78	I S F V L G F Y V T L V V S R W W S Q Y	97
301	CGAGAACTTGCCGTGGCCCCGACCGCCTCATGATCCAGGTGTCTAGCTTCGTGGAGGGCAA	360
98	E N L P W P D R L M I Q V S S F V E G K	117
361	GGATGAGGAAGGCCGTTTGCTGCGGCGCACGCTCATCCGCTACGCCATCCTGGGCCAAGT	420
118	D E E G R L L R R T L I R Y A I L G Q V	137
421	GCTCATCCTGCGCAGCATCAGCACCTCGGTCTACAAGCGCTTTCCCACTCTTCACCACCT	480
138	L I L R S I S T S V Y K R F P T L H H L	157
481	GGTGCTAGCAGGTTTTATGACCCATGGGGAACATAAGCAGTTGCAGAAGTTGGGCCTACC	540
158	V L A G F M T H G E H K Q L Q K L G L P	177
541	ACACAACACATTCTGGGTGCCCTGGGTGTGGTTTGCCAACTTGTCATGAAGGCCTATCT	600
178	H N T F W V P W V W F A N L S M K A Y L	197
601	TGGAGGTGGAATCCGGGACACCGTCTCTCCAGAGCCTGATGAATGAGGTGTGTACTTT	660
198	G G R I R D T V L L Q S L M N E V C T L	217
661	GCGTACTCAGTGTGGACAGCTGTATGCCTACGACTGGATAAGTATCCCATTTGGTGTAAC	720
218	R T Q C G Q L Y A Y D W I S I P L V Y T	237
721	ACAGGTGGTGACAGTGGCAGTATACAGCTTTTTCCTTGCTGCTTGATCGGGAGGCAGTT	780
238	Q V V T V A V Y S F F L A C L I G R Q F	257

FIGURE 8B

781	TCTGAACCCAAACAAGGACTACCCAGGCCATGAGATGGATCTGGTTGTGCCTGTCTTCAC	840
258	L N P N K D Y P G H E M D L V V P V F T	277
841	AATCCTGCAATTCTTATTCTACATGGGCTGGCTGAAGGTGGCAGAACAGCTCATCAACCC	900
278	I L Q F L F Y M G W L K V A E Q L I N P	297
901	CTTCGGGGAGGACGATGATGATTTTGAGACTAACTGGATCATTGACAGAAACCTGCAGGT	960
298	F G E D D D D F E T N W I I D R N L Q V	317
961	GTCCCTGTTGTCCGTGGATGGGATGCACCAGAACTTGCCTCCCATGGAACGTGACATGTA	1020
318	S L L S V D G M H Q N L P P M E R D M Y	337
1021	CTGGAACGAGGCAGCGCCTCAGCCGCCCTACACAGCTGCTTCTGCCAGGTCTCGCCGGCA	1080
338	W N E A A P Q P P Y T A A S A R S R R H	357
1081	TTCCTTCATGGGCTCCACCTTCAACATCAGCCTAAAGAAAGAAGACTTAGAGCTTTGGTC	1140
358	S F M G S T F N I S L K K E D L E L W S	377
1141	AAAAGAGGAGGCTGACACGGATAAGAAAGAGAGTGGCTATAGCAGCACCATAGGCTGCTT	1200
378	K E E A D T D K K E S G Y S S T I G C F	397
1201	CTTAGGACTGCAACCCAAAACTACCATCTTCCCTTGAAAGACTTAAAGACCAAACTATT	1260
398	L G L Q P K N Y H L P L K D L K T K L L	417
1261	GTGTTCTAAGAACCCCTCCTCGAAGGCCAGTGTAAGGATGCCAACCAGAAAAACCAGAA	1320
418	C S K N P L L E G Q C K D A N Q K N Q K	437
1321	AGATGTCTGGAAATTTAAGGGTCTGGACTTCTTGAAATGTGTTCCAAGGTTTAAGAGGAG	1380
438	D V W K F K G L D F L K C V P R F K R R	457
1381	AGGCTCCCATTTGTGGCCCAAGGCACCCAGCAGCCACCCTACTGAGCAGTCAGCACCCCTC	1440
458	G S H C G P Q A P S S H P T E Q S A P S	477
1441	CAGTTCAGACACAGGTGATGGGCCTTCCACAGATTACCAAGAAATCTGTCACATGAAAAA	1500
478	S S D T G D G P S T D Y Q E I C H M K K	497
1501	GAAACTGTGGAGTTTAACCTGAACATTCCAGAGAGCCCCACAGAACATCTTCAACAGCG	1560
498	K T V E F N L N I P E S P T E H L Q Q R	517
1561	CCGTTTGGACCAGATGTCAACCAATATACAGGCTCTAATGAAGGAGCATGCAGAGTCCTA	1620

FIGURE 8C

518 R L D Q M S T N I Q A L M K E H A E S Y 537

1621 TCCCTACAGGGATGAAGCTGGCACCAAACCTGTTCTCTATGAGTGATGCCTCACAGCCTG 1680
538 P Y R D E A G T K P V L Y E 551

1681 GCCCTGACTTGCAAGGATGCCAGCAGGGCACTGACCCAGTCAAAGGCACACAAGCAGCG 1740

1741 ACACCCAGGAGTGTGTTCCCACGACAGTCTAGCATGTAACTCAGAACCAAGAGTACTTAA 1800

1801 TAGTCCTGCCTGAAAACACCTGTATTTTACGATCTTTCCCAAACCTAAGGAGTTTAAATAAA 1860

1861 CGTGAATATTTCTTTTAGGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1916

FIGURE 9A

!!AA_MULTIPLE_ALIGNMENT 1.0
 PileUp of: *]

Symbol comparison table: GenRunData:blosom62.cmp CompCheck: 6430

GapWeight: 12
 GapLengthWeight: 4

pileup.msf MSF: 596 Type: P October 1, 1998 10:43 Check: 124 ..

Name: Human Len: 596 Check: 3272 Weight: 1.00
 Name: MouseBestrophin] Len: 596 Check: 6852 Weight: 1.00

	1		50
Human	MTITYTSQVA NARLGFSFRL LLCWRGSIYK LLYGEFLIFL LCYYIIRFIY		
MouseBestrophin]	MTITYTNKVA NARLGSFSSL LLCWRGSIYK LLYGEFLVFI FLYYSIRGLY		
	51		100
Human	RLALTEEQQL MFEKLTLYCD SYIQLIPISF VLGFYVTLVV TRWWNQYENL		
MouseBestrophin]	RMVLSSDQQL LFEKLALYCD SYIQLIPISF VLGFYVTLVV SRWWSQYENL		
	101		150
Human	FWPDRMLSLV SGFVEGKDEQ GRLLRRTLIR YANLGNVLIL RSVSTAVYKR		
MouseBestrophin]	PWPDRMLIQV SSFVEGKDEE GRLLRRTLIR YAILGQVLIL RSISTSIVYKR		
	151		200
Human	FPSAQHLVQA GFMTPAEHKQ LEKLSLPHNM FWVPWWVFAN LSMKAWLGGR		
MouseBestrophin]	FPTLHHLVLA GFMTGHEHKQ LQKLGLPHNT FWVPWWVFAN LSMKAYLGGR		
	201		250
Human	IRDPILLQSL LNECNTLRTQ CGHLYAYDWI SIPLVYTQVV TVAVYSFFLT		
MouseBestrophin]	IRDTVLLQSL MNEVCTLRTQ CGQLYAYDWI SIPLVYTQVV TVAVYSFFLA		
	251		300
Human	CLVGRQFLNP AKAYPGHELD LVVPVFTFLQ FFFYVGWLKV AEQLINPFGE		
MouseBestrophin]	CLIGRQFLNP NKDYPGHEMD LVVPVFTILO FLFYMGWLKV AEQLINPFGE		
	301		350
Human	DDDDFETNWI VDRNLQVSLL AVDEMHQDLP RMEPDYWNK PEPQPPYTAA		
MouseBestrophin]	DDDDFETNWI IDRNLQVSLL SVDGMHQNLP PMERDYMWNK AAPQPPYTAA		
	351		400
Human	SAQFRRASFV GSTFNISLTK EEMEFQPNQEDEEDAH AGIIGRFLGL		
MouseBestrophin]	SARSRRHSFM GSTFNISLTK EDLELWSKEE ADTDKESGY SSTIGCFLGL		
	401		450
Human	QSHDHHPPRA NSRTKLLWPK RESLLHEGLP KNHKAQKQV RGQEDNKAWK		
MouseBestrophin]	QPKNYHLPLK DLKTKLLCSK NPLL..EGQC KD.....ANQ KNQKD..VWK		
	451		500
Human	LKAVIDAFKSA PLYQRPQYYS APQTPLSPTP MFFPLEPSAP SKLHSVTGID		
MouseBestrophin]	FKGLDFLKCV PRFKRRGSHC GPQAPSS... ..HPTEQSAP SS...SDTG..		

FIGURE 9B

	501		550
Human	TKDKSLKTVS	SGAKKSFELL	SESDGALMEH PEVSQVRRKT VEFNLDMPE
MouseBestrophin]DGPSTDY QEICHMKKKT VEFNL.NIPE
	551		596
Human	IPENHLKE.P	LEQSPTNIHT	TLKDHDMPYW ALENRDEAHS -----
MouseBestrophin]	SPTEHLQRR	LDMSTNIQA	LMKEHAESY. ..PYRDEAGT KPVLYE

FIGURE 10

